Published in final edited form as:

Comp Biochem Physiol Part D Genomics Proteomics. 2013 September ; 8(3): 165–193. doi:10.1016/j.cbpd.2013.04.002.

Prediction of the protein components of a putative *Calanus finmarchicus* (Crustacea, Copepoda) circadian signaling system using a *de novo* assembled transcriptome

Andrew E. Christie*, Tiana M. Fontanilla, Katherine T. Nesbit, and Petra H. Lenz

Békésy Laboratory of Neurobiology, Pacific Biosciences Research Center, University of Hawaii at Manoa, 1993 East-West Road, Honolulu, HI 96822 USA

Abstract

Diel vertical migration and seasonal diapause are critical life history events for the copepod *Calanus finmarchicus*. While much is known about these behaviors phenomenologically, little is known about their molecular underpinnings. Recent studies in insects suggest that some circadian genes/proteins also contribute to the establishment of seasonal diapause. Thus, it is possible that in *Calanus* these distinct timing regimes share some genetic components. To begin to address this possibility, we used the well-established *Drosophila melanogaster* circadian system as a reference for mining clock transcripts from a 200,000+ sequence *Calanus* transcriptome; the proteins encoded by the identified transcripts were also deduced and characterized. Sequences encoding homologs of the *Drosophila* core clock proteins CLOCK, CYCLE, PERIOD and TIMELESS were identified, as was one encoding CRYPTOCHROME 2, a core clock protein in ancestral insect systems, but absent in *Drosophila*. *Calanus* transcripts encoding proteins known to modulate the *Drosophila* core clock were also identified and characterized, *e.g.* CLOCKWORK ORANGE, DOUBLETIME, SHAGGY and VRILLE. Alignment and structural analyses of the deduced *Calanus* proteins with their *Drosophila* counterparts revealed extensive sequence conservation, particularly in functional domains. Interestingly, reverse BLAST analyses of these sequences against all arthropod proteins typically revealed non-*Drosophila* isoforms to be most similar to the *Calanus* queries. This, in combination with the presence of both CRYPTOCHROME 1 (a clock input pathway protein) and CRYPTOCHROME 2 in *Calanus*, suggests that the organization of the copepod circadian system is an ancestral one, more similar to that of insects like *Danaus plexippus* than to that of *Drosophila*.

Keywords

circadian rhythm; diapause; photoperiodism; Illumina sequencing; functional genomics

1. Introduction

The precise timing of biological processes is a key component of physiological and behavioral control in all organisms. To date, a myriad of timing regimes have been

© 2013 Elsevier Inc. All rights reserved.

*Correspondence to: Dr. Andrew E. Christie, Békésy Laboratory of Neurobiology, Pacific Biosciences Research Center, University of Hawaii at Manoa, 1993 East-West Road, Honolulu, Hawaii 96822 USA, Phone: 808-956-5212; Fax: 808-956-6984, crabman@pbrc.hawaii.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

documented, including many operating on different time courses within a single species. The presence of multiple “clock” systems that operate with different timing cycles within a species, or a single individual, raises a fundamental question for the field of chronobiology: “do the molecular cascades that give rise to distinct timing regimes utilize common or unique genetic machinery to affect physiological/behavioral output?”. While the timing regimes that have been documented are extensive, circadian rhythms, physiological and behavioral patterns that oscillate with a period of approximately 24-hours and are synchronized to the solar day, are without question the best understood, at least at the molecular level.

The circadian system of the fruit fly *Drosophila melanogaster* is arguably the best understood in the animal kingdom (for a detailed review of the molecular underpinning of the *D. melanogaster* circadian system see Allada and Chung, 2010). In *Drosophila*, a large number of genes/proteins are known to play roles in the generation of circadian rhythmicity, including core clock proteins, input pathway proteins, output pathway proteins, which are responsible for time keeping, synchronizing the clock system to the environment, and transmitting the timing information to affect physiology/behavior, respectively. In brief, the core clock of *D. melanogaster* consists of an interactive feedback loop involving CLOCK (CLK), CYCLE (CYC), PERIOD (PER) and TIMELESS (TIM) proteins and genes. In *D. melanogaster*, the core clock is achieved by a CLK/CYC heterodimer binding to E-box elements in the promoters of the *per* and *tim* genes, resulting in their transcriptional activation late in the solar day. As a result, PER and TIM proteins are produced early in the evening, accumulating and forming heterodimers in the cytoplasm. At approximately midnight, the PER/TIM complex is translocated to the nucleus, where it interacts with the CLK/CYC heterodimer, inhibiting the latter’s ability to bind to DNA, and hence its ability to bind to and activate the *per* and *tim* genes. A number of other genes/proteins are critical for modulating the CLK/CYC/PER/TIM feedback loop, for example kinases that phosphorylate these proteins (*e.g.* DOUBLETIME [DBT] and SHAGGY [SGG]) and ligases that degrade the phosphorylated substrates (*e.g.* SUPERNUMERARY LIMBS [SLIMB]). In addition, the CLK/CYC heterodimer participates in additional feedback loops that have been shown to play critical roles in setting the phase and amplitude of the core molecular clock, as well as its rhythmic output. For example the complex is responsible for the transcriptional activation of the *par domain protein 1* (*pdp1*) and *vrille* (*vri*) genes, whose protein products activate and repress the *clk* and *cryptochrome 1* (*cry1*) genes, respectively; the blue light photoreceptor CRY1 is an important input pathway protein that promotes TIM degradation (for review see Allada and Chung, 2010).

Interestingly, while the circadian system of *Drosophila* is arguably the best understood in the animal kingdom, evolutionary studies of CRY duplication and loss in other members of the Insecta suggest that it is not a stereotypical system. Specifically, it is becoming increasingly apparent that the most ancestral organization of circadian systems in insects is one with two CRY genes/proteins, where CRY1 serves as a photic entrainer and CRY2 participates in the core clock as a repressor of CLK/CYC-mediated gene transcription (*e.g.* Zhu et al., 2005; Yuan et al., 2007; Zhu et al., 2008); *D. melanogaster* possesses *cry1*, but *cry2* appears to have been lost in this species (Yuan et al., 2007).

Since the late-1930s, it has been hypothesized that the genes/proteins involved in circadian clock systems are not only responsible for the generation of daily rhythms, but also are important contributors to seasonal photoperiodism (Bünning, 1936), the biological rhythm responsible for timing seasonal diapause. While this hypothesis was initially based solely on the fact that both rhythms track environmental day-night cycles, there currently is a large volume of experimental data from organisms as diverse as fungi and mammals in support of it, though perhaps not universally (Nelson et al., 2010).

The vast majority of studies linking the circadian and photoperiodism systems are from insects. In these animals, a variety of mechanisms have been proposed by which photoperiodic timing using circadian systems might be achieved (*e.g.* Saunders, 2010; Saunders and Bertossa, 2011; Saunders, 2013). For example, in the wasp *Nasonia vitripennis* it is hypothesized that the sensing of phase shifts in a pair of circadian pacemakers, one phase set by dawn and the other by dusk, are used to track seasonal changes in photoperiod (*e.g.* Saunders, 1974). Given the diversity of species in which the circadian and photoperiodism systems are linked, it is assumed that the relationship is an ancient one (Nelson et al., 2010). The diversity of mechanisms by which the two systems interact in insects is thought to result from gene loss or gain within the circadian systems of these animals, *e.g.* the CRY and TIM genes, as well as the extensive speciation that has occurred within this arthropod taxon (Saunders and Bertossa, 2011).

Like the insects, crustaceans are members of the Arthropoda, with the two subphyla now generally recognized as forming a clade, the Pancrustacea (Giribet and Edgecombe, 2012). While many crustaceans have been shown to exhibit robust circadian rhythms in physiology and behavior (de la Iglesia and Hsu, 2010; Farca Luna et al., 2010; Strauss and Dirksen, 2010), little is currently known about the genes/proteins or molecular cascades that underlie circadian rhythmicity in these animals, though their close phylogenetic relationship with the insects would predict similarities between them. Likewise, there is currently no information as to whether the circadian systems of crustaceans participate in the generation of other timekeeping systems, *e.g.* tidal and/or seasonal clocks, though it seems plausible that, as in insects, at least some components of the systems may be shared. The copepod crustacean *Calanus finmarchicus* provides a model in which to begin investigating the possibility of interactions between the circadian and seasonal photoperiodism systems in the Crustacea, in that it exhibits both a diel vertical migration within the water column (*e.g.* Hays, 2003; Rabindranath et al., 2011) and a seasonal subadult diapause (*e.g.* Hirche, 1996), behaviors that are likely to be under circadian and seasonal photoperiodic control, respectively. Moreover, *C. finmarchicus* has been the subject of a recent large-scale transcriptome sequencing project (Christie et al., 2013a), which now provides a rich resource for gene/protein discovery in this species. As a first step toward investigating the molecular mechanisms underlying biological rhythms in *Calanus*, we have mined its transcriptome for homologs of insect proteins known to play roles in the establishment of circadian clock systems.

2. Materials and methods

2.1. *De novo* transcriptome assembly

A *de novo* transcriptome for *C. finmarchicus* was generated as described in detail in a previous publication (Christie et al., 2013a). In brief, multiplexed gene libraries were prepared from RNA extracted from six developmental stages of wild-caught or laboratory cultured *C. finmarchicus*. The developmental stages included in the libraries were egg (which represents a mixture of embryonic stages), early nauplius (stages NI and NII), late nauplius (stages NV and NVI), early copepodite (stages CI and CII), late copepodite (stage CV) and adult female; all libraries were sequenced at the HudsonAlpha Institute for Biotechnology (Huntsville, AL, USA) in a single lane using an Illumina HiSeq 2000 instrument (Illumina Inc.). In total, 415,469,690,000 raw, 100-bp, paired-end reads were obtained. After quality filtering and trimming, the combined raw reads from all six developmental stage libraries were assembled *de novo* using Trinity 2012-03-17-IU_ZIH_TUNED software on a node of the National Center for Genome Analysis Support's (NCGAS; Indiana University, Bloomington, IN, USA) Mason Linux cluster. In total, 206,041 unique nucleotide sequences were generated using Trinity.

2.2. Transcriptome mining

Searches of the transcriptome assembly produced by Trinity were conducted using the DeCypher Tera-BLASTP algorithm on the TimeLogic DeCypher server located at Mount Desert Biological Laboratory (Salisbury Cove, ME, USA). For all searches, the DeCypher program database was set to the combined Trinity assembly, and the sequence of a known insect circadian protein, primarily ones from *D. melanogaster*, was used as the query. The default parameters for Tera-BLASTP were used for all searches. All hits were translated and checked manually for homology to the target query. The BLAST-generated E-value for each hit identified as encoding a putative target homolog is provided in Table 1, as is the length of each of the *Calanus* transcripts.

2.3. Analyses of protein conservation and structure

Analyses of protein conservation and structure were conducted using a protocol described in detail in a number of recent publications (e.g. Tilden et al., 2011; McCoole et al., 2011; McCoole et al., 2012a; McCoole et al., 2012b; Christie et al., 2013a; Christie et al., 2013b). To determine the known protein most similar to each of the *C. finmarchicus* circadian proteins identified in this study, the deduced *Calanus* sequence was used to query the annotated proteins dataset present in FlyBase (version FB2013_01), as well as the non-redundant arthropod protein dataset (taxid:6656) curated in GenBank (excluding *C. finmarchicus* proteins, obvious partial proteins, synthetic constructs, and provisional/unannotated hypothetical protein sequences); for both searches the blastp algorithm was used (Altschul et al., 1997). Table 2 summarizes the results of these reverse blastp searches.

To determine amino acid identity/similarity between proteins and structural motifs (see below), the sequences in question were aligned using MAFFT version 7 (<http://align.bmr.kyushu-u.ac.jp/mafft/online/server/>; Katoh et al., 2002; Katoh and Toh, 2008), and amino acid identity/similarity was subsequently determined using the alignment output. Specifically, percent identity was calculated as the number of identical amino acids (denoted by “*” in the MAFFT output) divided by the total number of amino acids in the longest sequence (x100), while amino acid similarity was calculated as the number of identical and similar amino acids (the latter denoted by the “:” and “.” symbols in the protein alignment) divided by the total number of amino acids in longest sequence (x100).

Protein structural motifs were analyzed using the online program SMART (<http://smart.embl-heidelberg.de/>; Schultz et al., 1998; Letunic et al., 2009); the default parameters of the program were used for all analyses. A common highlighting scheme has been used to denote functional domains in all figures: basic region leucine zipper, pink; casein kinase II regulatory subunit domain, black; coiled coil region, gray; D domain of beta-TrCP, blue; FBOX, dark blue; helix-loop-helix, yellow; orange domain, red; PAC, dark green; PAS, light green; protein phosphatase 2A catalytic domain, dark red; serine/threonine protein kinase catalytic domain, purple; WD40, teal.

2.4. Vetting of deduced protein sequences using publicly accessible expressed sequence tags

To vet the quality of the protein sequences deduced from the Trinity assembly, each of the predicted proteins was used to query the extant *C. finmarchicus* ESTs (11,461 in total as of January 1, 2013) curated at NCBI for identical/highly similar sequences (Lenz et al., 2012). Specifically, searches for *Calanus* ESTs encoding proteins corresponding to those deduced from Trinity assembled sequences were conducted using the tblastn algorithm (protein query vs. translated nucleic database) on NCBI. Here, the database was set to “Expressed sequence tags (est)” and limited to *C. finmarchicus* sequences (taxid:6837). The default parameters of tblastn were used for all searches.

3. Results

3.1. *In silico* identification of putative *Calanus finmarchicus* circadian proteins

It is generally agreed that all clock systems can be divided into three subunits: the clock, input pathways to the clock, and output pathways from the clock, each of which consists of multiple gene/protein components (Allada and Chung, 2010). Furthermore, the genes/proteins that form the clock can be subdivided into those forming the core clock and those involved in its modulation, *i.e.* clock-associated proteins. The results of transcriptome searches and protein predictions/analyses are presented in alphabetical order according to their theoretical role within the putative *C. finmarchicus* circadian system, *i.e.* core clock proteins, clock-associated proteins or input pathway proteins. The output pathway transcripts/proteins putatively used by the *Calanus* circadian system, specifically peptide hormone precursors and their receptors, were the subject of a separate study, and readers are directed to Christie et al. (2013a) for descriptions and discussions of these data. It should be noted that the protein sequences presented were deduced from a *de novo* transcriptome assembly, and as such, should be taken as theoretical rather than biochemically-confirmed.

3.1.1. Core clock proteins

3.1.1.1. CLOCK (CLK): Using the amino acid sequence of *D. melanogaster* CLOCK (Accession No. **AAC62234**; Bae et al., 1998), one transcript was identified as encoding a putative *C. finmarchicus* CLOCK (Calfi-CLK; Table 1). The protein deduced from this transcript is 540 amino acids in length, and appears to be an amino (N)-terminal partial sequence. Alignment of the *Drosophila* and *Calanus* proteins is shown in Figure 1. As can be seen from this alignment, the extant portion of Calfi-CLK shares 43.9% identity and 76.3% similarity with its *Drosophila* counterpart.

Structural analysis of Calfi-CLK using the online program SMART identified one helix-loop-helix domain, two PAS domains and one PAC domain within the extant sequence (Fig. 1). An identical complement of helix-loop-helix, PAS, and PAC domains are predicted by SMART for the *Drosophila* protein, as were three coiled coil regions (Fig. 1). The portions of the *Calanus* and *Drosophila* proteins that comprise the helix-loop-helix, PAS, and PAC domains show higher levels of amino acid conservation than do the two proteins when compared as wholes (helix-loop-helix, 68.6% identity/92.2% similarity; PAS-1, 44.8% identity/82.1% similarity; PAS-2, 65.7% identity/97.0% similarity; PAC, 77.3% identity/95.5% similarity).

The data just presented strongly suggest that Calfi-CLK is a true member of the CLOCK protein family. However, to strengthen our confidence in this hypothesis, reverse BLAST analyses of Calfi-CLK against both the proteins annotated in FlyBase and the non-redundant arthropod protein dataset present in GenBank was undertaken. For the former search, CLOCK, isoform A, (FlyBase No. **FBpp0076500**; Adams et al., 2000) was identified as the *Drosophila* protein most similar to that of Calfi-CLK (Table 2). For the latter search, an isoform of CLOCK from the firebrat *Thermobia domestica* (Accession No. **BAJ16353**; Kamae et al., 2010) was identified as most similar to the *Calanus* query (Table 2). These data too are strongly supportive of Calfi-CLK being a true CLOCK family member.

3.1.1.2. CRYPTOCHROME 2 (CRY2): Using the amino acid sequence of monarch butterfly *Danaus plexippus* CRYPTOCHROME 2 (Accession No. **ABA62409**; Zhu et al., 2005), one *C. finmarchicus* transcript was identified as encoding a putative CRYPTOCHROME 2 (Calfi-CRY2; Table 1). The protein deduced from this transcript is 794 amino acids in length and appears to be a full-length protein. Figure 2 shows the alignment of the *Danaus* and *Calanus* proteins, and as can be seen here, Calfi-CRY2 shares

52.5% identity and 74.8% similarity with the query sequence. No functional domains were identified in Calfi-CRY2 by SMART, nor were any identified in the *Danaus* protein using this online program. As *D. melanogaster* does not possess a *cry2* gene (e.g. Tomioka and Matsumoto, 2010), Calfi-CRY2 was BLASTed first against the *D. plexippus* proteins present in GenBank, and here, the same protein used as the original query, i.e. CRYPTOCHROME 2, was found to be most similar to the *Calanus* protein (Table 2). When the database was expanded to all arthropod sequences, CRYPTOCHROME 2 from the cockroach *Rhyparobia maderae* (Accession No. **AGA01579**; Werckenthin et al., 2012) was identified as the protein most similar to Calfi-CRY2 (Table 2). These data, in conjunction with those presented above, strongly support Calfi-CRY2 being a member of the CRYPTOCHROME 2 family.

3.1.1.3. CYCLE (CYC): Using the amino acid sequence of *D. melanogaster* CYCLE (Accession No. **AAF49107**; Adams et al., 2000), one transcript was identified as encoding a putative *C. finmarchicus* CYCLE (Calfi-CYC; Table 1). The protein deduced from this transcript is 685 amino acids in length, and appears to be a full-length protein. Alignment of the *Drosophila* and *Calanus* proteins is shown in Figure 3. As can be seen from this alignment, Calfi-CYC shares 31.2% identity and 50.1% similarity with its *Drosophila* counterpart.

Structural analysis of Calfi-CYC using the online program SMART identified one helix-loop-helix domain, two PAS domains and one PAC domain within this protein (Fig. 3). An identical complement of similarly sized and placed helix-loop-helix, PAS, and PAC domains were predicted by SMART for the *Drosophila* query (Fig. 3). The helix-loop-helix, PAS, and PAC domains of the *Calanus* and *Drosophila* proteins share higher levels of amino acid conservation than do the two proteins when compared as wholes (helix-loop-helix, 73.2% identity/89.3% similarity; PAS-1, 60.3% identity/88.2% similarity; PAS-2, 39.7% identity/79.4% similarity; PAC, 44.4% identity/84.4% similarity).

The high levels of amino acid and functional domain conservation seen between Calfi-CYC and its *Drosophila* counterpart are strongly suggestive of the former protein being a member of the CYCLE family. To further strengthen this hypothesis, reverse BLAST analyses of the *Calanus* protein against the annotated proteins in FlyBase and the non-redundant arthropod proteins curated in GenBank was undertaken. The query against FlyBase revealed CYCLE, isoform A, (FlyBase No. **FBpp0074693**; Adams et al., 2000) to be the *Drosophila* protein most similar to that of Calfi-CYC (Table 2), while the GenBank search revealed an isoform of CYCLE from the beetle *Tribolium castaneum* (Accession No. **ABO86538**; Zhu and Reppert, unpublished submission) to be most similar to the *Calanus* query (Table 2). These data reinforce the hypothesis that Calfi-CYC is a true CYCLE family member.

3.1.1.4. PERIOD (PER): Using the amino acid sequence of a *D. melanogaster* PERIOD isoform (Accession No. **AAF45804**; Adams et al., 2000), two transcripts were identified as encoding putative *C. finmarchicus* PERIOD proteins (Calfi-PER I and II; Table 1). Both Calfi-PER I and Calfi-PER II appear to be full-length proteins, with the former being 1324 amino acids in length, and the latter, 1581 amino acids long. Alignment of the *Drosophila* query with each of the *Calanus* proteins is shown in Figure 4A. As can be seen from these alignment, Calfi-PER I shares 25.3% identity and 54.5% similarity with its *Drosophila* counterpart (Fig. 4A1), while Calfi-PER II shares 22.0% identity and 49.3% similarity with the *D. melanogaster* protein (Fig. 4A2). Alignment of the two *Calanus* proteins shows that they share 36.2% identity and 62.7% similarity with one another (Fig. 4B).

Structural analyses of the Calfi-PERs using the online program SMART identified a pair of PAS domains in each protein (Fig. 4). In addition, a single PAC domain was predicted in

Calfi-PER II, but not Calfi-PER I (Fig. 4). Two PAS and one PAC domains were predicted by SMART for the *Drosophila* query (Fig. 4A). Comparisons of the PAS domain portions of Calfi-PER I and II with those identified in the *Drosophila* query show higher levels of amino acid conservation than do either of the two *Calanus* proteins and the *Drosophila* protein when compared in their entirety (*Drosophila* vs. Calfi-PER I: PAS-1, 50.7% identity/83.1% similarity and PAS-2, 37.7% identity/69.6% similarity; *Drosophila* vs. Calfi-PER II: PAS-1, 42.9% identity/76.6% similarity and PAS-2, 30.1% identity/69.9% similarity). Likewise the PAC domain portions of Calfi-PER II and the *Drosophila* query show greater amino acid conservation than do the two proteins when compared as complete sequences (*i.e.* 50.0% identity/81.8% similarity). Moreover, while no PAC domain was identified in Calfi-PER I, the *Drosophila* PAC region and the corresponding portion of Calfi-PER I show higher levels of amino acid composition (*i.e.* 61.4% identity/84.1% similarity) than do the PACs of *Drosophila* and Calfi-PER II. The portions of the *Calanus* PERs that form the two PAS and single PAC domains also show higher levels of amino acid identity/similarity than do the proteins when compared as wholes (*i.e.* PAS-1, 67.1%/91.4%; PAS-2, 55.7%/80.0%; PAC, 61.4%/90.9%).

The data just presented support Calfi-PER I and II being members of the PERIOD protein family. However, reverse BLAST analyses of Calfi-PERs against both the annotated proteins in FlyBase and the non-redundant arthropod protein dataset present in GenBank was undertaken to strengthen our confidence in this hypothesis. Searches of FlyBase using either of the *Calanus* proteins as the query identified PERIOD, isoform B, (FlyBase No. **FBpp0304590**) as the *Drosophila* protein most similar to both proteins (Table 2). Queries of the non-redundant arthropod sequences in GenBank identified *R. maderae* PERIOD (Accession No. **AGA01525**; Werckenthin et al., 2012) as the protein most similar to Calfi-PER I, and PERIOD, isoform 1, from the cricket *Apteronomobius asahinai* (Accession No. **BAL72155**; Takekata, Matsuura, Goto, Numata and Satoh, unpublished submission) as the most similar protein to Calfi-PER II. These data also strongly support the two Calfi-PERs being true members of the PERIOD family.

3.1.1.5. TIMELESS (TIM): Using the amino acid sequence of a *D. melanogaster* TIMELESS isoform (Accession No. **AAC46920**; Myers et al., 1995), one *C. finmarchicus* transcript was identified as encoding a putative TIMELESS (Calfi-TIM; Table 1). The protein deduced from this transcript is 1225 amino acids in length and appears to be a full-length protein. Figure 5 shows the alignment of the *Drosophila* and *Calanus* proteins, and as can be seen here Calfi-TIM shares 31.7% identity and 63.9% similarity with the query sequence. No functional domains were identified in Calfi-TIM by SMART, nor were any identified in the *Drosophila* query using this online program.

BLASTs of Calfi-TIM against the annotated proteins in FlyBase identified TIMELESS, isoform Q, (FlyBase No. **FBpp0305455**) as the most similar *D. melanogaster* protein to the query sequence, while reverse BLAST analysis against all non-redundant arthropod proteins in GenBank identified TIMELESS from the cricket *Gryllus bimaculatus* (Accession No. **BAJ16356**; Uryu and Tomioka, 2010) as the protein most similar to Calfi-TIM (Table 2). These data, in conjunction with those presented above, are strongly supportive of Calfi-TIM being a member of the TIMELESS family.

3.1.2. Clock-associated proteins

3.1.2.1. Casein Kinase II (CKII)

CKII α -subunit: Using the amino acid sequence of an isoform of *D. melanogaster* CASEIN KINASE II α -subunit (Accession No. **AAN11415**; Adams et al., 2000), one *C. finmarchicus* transcript was identified as encoding a putative CASEIN KINASE II α (Calfi-CKII α ; Table

1). The protein deduced from this transcript is 370 amino acids in length and appears to be a full-length protein. Figure 6A shows the alignment of the *Drosophila* and *Calanus* proteins, and as can be seen here, Calfi-CKII α shares 79.5% identity and 87.6% similarity with the query sequence. Structural analyses of Calfi-CKII α and the *Drosophila* query using the online program SMART identified a single serine/threonine protein kinase catalytic domain in each protein (Fig. 6A). The serine/threonine protein kinase catalytic domains of the *Calanus* and *Drosophila* proteins are nearly identical in sequence, sharing 92.0% amino acid identity/96.9% amino acid similarity. While the amino acid and structural domain conservation just described is supportive of Calfi-CKII α being a member of the CASEIN KINASE II α subunit family, reverse BLAST analyses of this sequence against both the proteins annotated in FlyBase and the non-redundant arthropod protein dataset present in GenBank was undertaken to help confirm this hypothesis. The search of FlyBase using Calfi-CKII α as the query, identified CASEIN KINASE II α , isoform A, (FlyBase No. **FBpp0070041**; Adams et al., 2000) as the *Drosophila* protein most similar to the *Calanus* query (Table 2); this protein is the same as was used for the initial identification of Calfi-CKII α . Querying of the non-redundant arthropod sequences in GenBank identified CASEIN KINASE II subunit α from the ant *Camponotus floridanus* (Accession No. **EFN62439**; Bonasio et al., 2010) as the protein most similar to Calfi-CKII α . These data too are strongly supportive of Calfi-CKII α being a true member of the CASEIN KINASE II α -subunit family.

CKII β -subunit: Using the amino acid sequence of an isoform of *D. melanogaster* CASEIN KINASE II β -subunit (Accession No. **AAF48093**; Adams et al., 2000), one *C finmarchicus* transcript was identified as encoding a putative CASEIN KINASE II β (Calfi-CKII β ; Table 1). The protein deduced from this transcript is 216 amino acids in length and appears to be a full-length protein. Figure 6B shows the alignment of the *Drosophila* and *Calanus* proteins, and as can be seen here Calfi-CKII β shares 89.4% identity and 96.3% similarity with the query sequence.

Structural analyses of Calfi-CKII β and the *Drosophila* query using the online program SMART identified a single casein kinase II regulatory subunit domain in each protein (Fig. 6B). The casein kinase II regulatory subunit domains of the *Calanus* and *Drosophila* proteins are essentially identical in sequence, exhibiting 92.4% identity/98.4% similarity in amino acid composition.

The amino acid and structural domain conservation just described is supportive of Calfi-CKII β being a member of the CASEIN KINASE II β family. However, to strengthen our confidence in this hypothesis, reverse BLAST analyses of this sequence against both the proteins annotated in FlyBase and the non-redundant arthropod protein dataset present in GenBank was undertaken. The search of FlyBase using Calfi-CKII β as the query, identified CASEIN KINASE II β , isoform K, (FlyBase No. **FBpp0305406**) as the *Drosophila* protein most similar to the *Calanus* query (Table 2). Querying of the non-redundant arthropod sequences in GenBank identified the same *Drosophila* protein used to initially identify Calfi-CKII β , *i.e.* *D. melanogaster* CASEIN KINASE II subunit β , isoform B, as the protein most similar to the *Calanus* query. These data too support Calfi-CKII β being a true member of the CASEIN KINASE II β -subunit family.

3.1.2.2. CLOCKWORK ORANGE (CWO): Using the amino acid sequence of a *D. melanogaster* CLOCKWORK ORANGE isoform (Accession No. **AAF54527**; Adams et al., 2000), one transcript was identified as encoding a putative *C finmarchicus* CLOCKWORK ORANGE (Calfi-CWO; Table 1). The protein deduced from this transcript is 587 amino acids in length, and appears to be a full-length protein. Alignment of the *Drosophila* and *Calanus* proteins is shown in Figure 7. As can be seen from this alignment, Calfi-CWO

shares 24.8% identity and 51.3% similarity with its *Drosophila* counterpart. Structural analyses of Calfi-CWO and the *Drosophila* query using the online program SMART identified one helix-loop-helix and one orange domain within each protein (Fig. 7). Comparisons of the amino acid sequences of the functional domains in the two proteins revealed them to share more identity/similarity than do Calfi-CWO and the *Drosophila* query when compared in their entirety (helix-loop-helix, 80.7%/96.5%; orange, 31.9%/70.2%).

While the overall sequence conservation between Calfi-CWO and its *Drosophila* counterpart is somewhat low, BLASTing of the *Calanus* protein against the annotated sequences in FlyBase identified the *Drosophila* query, CLOCKWORK ORANGE, isoform A, (FlyBase No. **FBpp0081723**) as the most similar *D. melanogaster* protein to it. Likewise, reverse BLAST analysis against all non-redundant arthropod proteins in GenBank identified a CLOCKWORK ORANGE-like homolog, Hairy/enhancer-of-split related with YRPW motif protein 2, from the ant *C. floridanus* (Accession No. **EFN67685**; Bonasio et al., 2010) as the protein most similar to Calfi-CWO (Table 2). These data, in conjunction with those presented earlier, provide support for Calfi-CWO being a true CLOCKWORK ORANGE family member.

3.1.2.3. DOUBLETIME (DBT): Using the amino acid sequence of a *D. melanogaster* isoform of discs overgrown/DOUBLETIME (Accession No. **AAF57110**; Adams et al., 2000), two transcripts were identified as encoding putative *C. finmarchicus* DOUBLETIME proteins (Calfi-DBT I and II; Table 1). Both *Calanus* sequences appear to be full-length proteins, with Calfi-DBT I being 312 amino acids in overall length, and Calfi-DBT II 635 amino acids long. Alignment of the *Drosophila* query with each *Calanus* DOUBLETIME is shown in Figure 8A. As the alignments show, Calfi-DBT I shares 58.6% identity and 67.7% similarity with the *Drosophila* query (Fig. 8A1), while Calfi-DBT II shares 38.1% identity and 54.6% similarity with the *D. melanogaster* protein (Fig. 8A2). Alignment of the Calfi-DBT I and II show that they share 38.7% identity and 46.5% similarity with one another (Fig. 8B).

Structural analyses of the *Calanus* and *Drosophila* DOUBLETIME proteins using the online program SMART identified one serine/threonine protein kinase catalytic domain in each protein (Fig. 8). Comparisons of the serine/threonine protein kinase catalytic domains of Calfi-DBT I and II with that identified in the *Drosophila* query show higher levels of amino acid conservation than do either of the two *Calanus* proteins and the *Drosophila* query when compared in their entirety (*Drosophila* vs. Calfi-DBT I, 84.2% identity/96.6% similarity; *Drosophila* vs. Calfi-DBT II, 73.2% identity/93.6% similarity). Likewise, the serine/threonine protein kinase catalytic domains of the *Calanus* DBTs show higher levels of amino acid identity/similarity than do the proteins when compared as wholes (*i.e.* 82.6%/96.2%).

The data just presented are supportive of Calfi-DBT I and II being members of the DOUBLETIME protein family. To further our confidence in this hypothesis, reverse BLAST analyses of these sequences against both the proteins annotated in FlyBase and the non-redundant arthropod protein dataset present in GenBank was undertaken. Searches of FlyBase using each of the *Calanus* proteins as the query, identified discs overgrown, isoform D, (FlyBase No. **FBpp0306615**) as the *Drosophila* protein most similar to both Calfi-DBT I and II (Table 2). Queries of the non-redundant arthropod sequences in GenBank identified a putative casein kinase 1 delta b from the tick *Rhipicephalus pulchellus* (Accession No. **JAA55893**; Tan, Francischetti, Manjunatha and Ribeiro, unpublished submission) as the protein most similar to each of the *Calanus* DBTs. These data too are strongly supportive of Calfi-DBT I and II being true members of the DOUBLETIME family.

3.1.2.4. PAR Domain Protein 1e (PDP1e): Searches of the *C. finmarchicus* transcriptome using the sequence of *D. melanogaster* PAR DOMAIN PROTEIN 1e (Accession No. **AAF04509**; Lin et al., 1997) failed to identify any hits that convincingly encoded PDP1e homologs (data not shown).

3.1.2.5. Protein phosphatase 1 (PP1): Using the amino acid sequence of an isoform of *D. melanogaster* PROTEIN PHOSPHATASE 1 (Accession No. **CAA39820**; Dombrádi et al., 1990), one *C. finmarchicus* transcript was identified as encoding a putative PROTEIN PHOSPHATASE 1 (Calfi-PP1; Table 1). The protein deduced from this transcript is 341 amino acids in length and appears to be a full-length protein. Figure 9 shows the alignment of the *Drosophila* and *Calanus* proteins, and as can be seen here Calfi-PP1 shares 86.5% identity and 94.4% similarity with the query sequence.

Structural analyses of Calfi-PP1 and the *Drosophila* query using the online program SMART identified a single protein phosphatase 2A catalytic domain in each protein (Fig. 9). Comparison of the sequences of the *Calanus* and *Drosophila* protein phosphatase 2A catalytic domains show that they share 95.6% identity/100% similarity in amino acid sequence.

The amino acid and structural domain conservation just described is supportive of Calfi-PP1 being a member of the PROTEIN PHOSPHATASE 1 family. However, reverse BLAST analyses of this sequence against both the proteins annotated in FlyBase and the non-redundant arthropod protein dataset present in GenBank was undertaken to strengthen our confidence in this hypothesis. The search of FlyBase using identified protein phosphatase 1alpha at 96A, isoform B, (FlyBase No. **FBpp0306442**) as the *Drosophila* protein most similar to Calfi-PP1 (Table 2). Querying of the non-redundant arthropod sequences in GenBank identified serine/threonine-protein phosphatase alpha-1 from the ant *C. floridanus*, (Accession No. **EFN69572**; Bonasio et al., 2010) as the protein most similar to the *Calanus* protein. These data further support Calfi-PP1 being a true member of the PROTEIN PHOSPHATASE 1 family.

3.1.2.6. PROTEIN PHOSPHATASE 2A (PP2A)

PP2A catalytic subunit – MICROTUBULE STAR (MTS): Using the amino acid sequence of a *D. melanogaster* isoform of the PP2A catalytic subunit MICROTUBULE STAR (Accession No. **AAF52567**; Adams et al., 2000), two transcripts were identified as encoding putative *C. finmarchicus* MICROTUBULE STAR proteins (Calfi-MTS I and II; Table 1). Both *Calanus* proteins are 311 amino acids long and appear to be full-length sequences. Alignment of the *Drosophila* query with each *Calanus* protein is shown in Figure 10A. As the alignments show, Calfi-MTS I shares 92.3% identity and 98.1% similarity with the *Drosophila* query (Fig. 10A1), while Calfi-MTS II shares 76.2% identity and 90.7% similarity with the *D. melanogaster* protein (Fig. 10A2); Calfi-MTS I and II share 76.5% identity and 92.3% similarity with one another (Fig. 10B).

Structural analyses of the *Calanus* and *Drosophila* MTS proteins using the online program SMART identified one protein phosphatase 2A catalytic domain in each protein (Fig. 10). Comparisons of the catalytic domains between the proteins show they are nearly identical in amino acid composition (*Drosophila* vs. Calfi-MTS I, 94.8% identity/100% similarity; *Drosophila* vs. Calfi-MTS II, 80.8% identity/94.5% similarity; Calfi-MTS I vs. Calfi-MTS II, 80.4% identity/94.8% similarity).

While the data just presented support Calfi-MTS I and II being members of the MICROTUBULE STAR family, reverse BLAST analyses of these sequences against both the proteins annotated in FlyBase and the non-redundant arthropod protein dataset present in

GenBank were undertaken to strengthen our confidence in this hypothesis. Searches of FlyBase using Calfi-MTS I or II as the query identified the same protein used for their initial identification, MICROTUBULE STAR (FlyBase No. **FBpp0079148**; Adams et al., 2000), as the *Drosophila* protein most similar to each protein (Table 2). Queries of the non-redundant arthropod sequences in GenBank identified serine/threonine-protein phosphatase PP-V from the louse *Pediculus humanus corporis* (Accession No. **EEB13988**; Kirkness et al., unpublished submission) as the protein most similar to each *Calanus* sequence. These data too are strongly supportive of Calfi-MTS I and II being true MICROTUBULE STAR family members.

PP2A regulatory subunit – WIDERBORST (WBT): Using the amino acid sequence of an isoform of the *D. melanogaster* PP2A regulatory subunit WIDERBORST (Accession No. **AAF56720**; Adams et al., 2000), two transcripts were identified as encoding putative *C. finmarchicus* WIDERBORST proteins (Calfi-WBT I and II; Table 1). Both *Calanus* WBTs appear to be full-length proteins, with Calfi-WBT I being 578 amino acids in overall length, and Calfi-WBT II 480 amino acids long.

Alignment of the *Drosophila* query with each of the *Calanus* proteins is shown in Figure 11A. As can be seen from these alignments, Calfi-WBT I shares 54.7% identity and 77.2% similarity with its *Drosophila* counterpart (Fig. 11A1), while Calfi-WBT II shares 75.2% identity and 84.7% similarity with the *D. melanogaster* protein (Fig. 11A2); the two *Calanus* proteins share 52.6% identity and 71.6% similarity with one another (Fig. 11B). No functional domains were identified by SMART in either of the *Calanus* WBTs, nor were any identified in the *Drosophila* query using this online program.

The data just presented suggest that the Calfi-WBT I and II are members of the WIDERBORST family. To further strengthen our confidence in this hypothesis, reverse BLAST analyses of each protein against both the proteins annotated in FlyBase and the non-redundant arthropod protein dataset present in GenBank was undertaken. The search of FlyBase using Calfi-WBT I as the query, identified protein phosphatase 2A regulatory B subunit, isoform Q, (FlyBase No. **FBpp0307570**) as the *Drosophila* protein most similar to it (Table 2), while the query using Calfi-WBT II returned WIDERBORST, isoform F, (FlyBase No. **FBpp0084575**) as the most similar *D. melanogaster* sequence (Table 2). Queries of the non-redundant arthropod sequences in GenBank identified serine/threonine-protein phosphatase 2A 56 kDa regulatory subunit delta isoform from the ant *C. floridanus* (Accession No. **EFN69797**; Bonasio et al., 2010) as the protein most similar to Calfi-WBT I, and a putative serine/threonine protein phosphatase 2a regulatory subunit from the tick *R. pulchellus* (Accession No. **JAA57743**; Tan, Francischetti, Manjunatha and Ribeiro, unpublished submission) as the most similar protein to Calfi-WBT II. These data too support Calfi-WBT I and II being members of the WIDERBORST family.

PP2A regulatory subunit – TWINS (TWS): Using the amino acid sequence of an isoform of the *D. melanogaster* PP2A regulatory subunit TWINS (Accession No. **AAF54498**; Adams et al., 2000), one transcript was identified as encoding a putative *C. finmarchicus* TWINS (Calfi-TWS; Table 1). The protein deduced from this transcript is 445 amino acids in length, and appears to be a full-length protein. Alignment of the *Drosophila* and *Calanus* proteins is shown in Figure 12. As can be seen from this alignment, Calfi-TWS shares 74.5% identity and 86.4% similarity with its *Drosophila* counterpart.

Structural analyses of Calfi-TWS and the *Drosophila* query using the online program SMART identified seven WD40 domains within each protein (Fig. 12); the portions of the two proteins that comprise these domains share higher levels of amino acid conservation than do the two proteins when compared as wholes (*i.e.* WD40-1, 90.7% identity/100%

similarity; WD40-2, 90.2% identity/100% similarity; WD40-3, 95.0% identity/97.5% similarity; WD40-4, 90.2% identity/100% similarity; WD40-5, 92.3% identity/100% similarity; WD40-6, 87.3% identity/100% similarity; WD40-7, 94.7% identity/100% similarity).

The data just presented support Calfi-TWS being a member of the TWINS family. However, reverse BLAST analyses of this protein against both the proteins annotated in FlyBase and the non-redundant arthropod protein dataset present in GenBank was undertaken to strengthen our confidence in this hypothesis. BLASTing of Calfi-TWS against the annotated sequences in FlyBase identified TWINS, isoform H, (FlyBase No. **FBpp0081671**) as the most similar *D. melanogaster* protein to the query sequence, while reverse BLAST analysis against all non-redundant arthropod proteins in GenBank identified protein phosphatase 2A regulatory subunit B from the crab *Scylla paramamosain* (Accession No. **AFK24473**; Liu and Ye, unpublished submission) as the protein most similar to Calfi-TWS (Table 2). These data, in conjunction with those presented above, provide strong support for Calfi-TWS being a member of the TWINS family.

3.1.2.7. SHAGGY (SGG): Using the amino acid sequence of a *D. melanogaster* SHAGGY isoform (Accession No. **AAN09084**; Adams et al., 2000), one transcript was identified as encoding a putative *C. finmarchicus* SHAGGY (Calfi-SGG; Table 1). The protein deduced from this transcript is 416 amino acids in length, and appears to be a full-length protein. Alignment of the *Drosophila* and *Calanus* proteins is shown in Figure 13. As can be seen from this alignment, Calfi-SGG shares 61.7% identity and 76.3% similarity with its *Drosophila* counterpart.

Structural analyses of Calfi-SGG and the *Drosophila* query using the online program SMART identified one serine/threonine protein kinase catalytic domain within each protein (Fig. 13); the amino acids that comprise these domains are highly conserved between the two proteins (*i.e.* 86.0% identity/97.9% similarity).

The high levels of amino acid and functional domain conservation seen between Calfi-SGG and its *Drosophila* counterpart are strongly suggestive of the former protein being a member of the SHAGGY family. To further strengthen our confidence this hypothesis, reverse BLAST analyses of the *Calanus* protein against the annotated proteins in FlyBase and the non-redundant arthropod proteins curated in GenBank was undertaken. The query against FlyBase revealed SHAGGY, isoform A, (FlyBase No. **FBpp0070450**; Adams et al., 2000) to be the *Drosophila* protein most similar to that of Calfi-SGG (Table 2), while the GenBank search of all arthropod proteins revealed glycogen synthase kinase 3 beta (SHAGGY) from the water flea *Daphnia pulex* (Accession No. **EFX72595**; Colbourne et al., 2011) to be the most similar protein to the *Calanus* query (Table 2). These data reinforce the hypothesis that Calfi-SGG is a true member of the SHAGGY family.

3.1.2.8. SUPERNUMERARY LIMBS (SLIMB): Using the amino acid sequence of an isoform of *D. melanogaster* SUPERNUMERARY LIMBS (Accession No. **AAF55853**; Adams et al., 2000), one transcript was identified as encoding a putative *C. finmarchicus* SUPERNUMERARY LIMBS (Calfi-SLIMB; Table 1). The protein deduced from this transcript is 551 amino acids in length, and appears to be a full-length sequence. Alignment of the *Drosophila* and *Calanus* proteins is shown in Figure 14. As can be seen from this alignment, Calfi-SLIMB shares 68.8% identity and 83.5% similarity with its *Drosophila* counterpart.

Structural analysis of Calfi-SLIMB using the online program SMART identified one D domain of beta-TrCP, one FBOX domain, and seven WD40 domains within the protein (Fig.

14); an identical complement of structural domains was predicted by SMART for the *Drosophila* sequence (Fig. 14). The portions of the two proteins that comprise the above mentioned functional domains show higher levels of amino acid conservation than do Calfi-SLIMB and its *Drosophila* counterpart when compared in their entirety (D domain of beta-TrCP, 57.5% identity/87.5% similarity; FBOX, 82.5% identity/97.5% similarity; WD40-1, 89.5% identity/94.7% similarity; WD40-2, 79.5% identity/87.2% similarity; WD40-3, 76.1% identity/78.3% similarity; WD40-4, 94.7% identity/100% similarity; WD40-5, 94.7% identity/100% similarity; WD40-6, 94.7% identity/100% similarity; WD40-7, 89.5% identity/100% similarity).

While the data presented above are strongly suggestive of Calfi-SLIMB being a true member of the SUPERNUMERARY LIMBS family, reverse BLAST analyses of Calfi-SLIMB against both the proteins annotated in FlyBase and the non-redundant arthropod protein dataset present in GenBank was undertaken. The search of FlyBase identified SLIMB, isoform C, (FlyBase No. **FBpp0306059**) as the *Drosophila* protein most similar to that of Calfi-SLIMB (Table 2), while F-box/WD repeat-containing protein 1A from the ant *Acromyrmex echinatio* (Accession No. **EG159076**; Nygaard and Zhang, unpublished submission) was identified as the most similar protein to the *Calanus* query using the GenBank dataset (Table 2). These data also strongly support Calfi-SLIMB being a true SUPERNUMERARY LIMBS family member.

3.1.2.9. VRILLE (VRI): Using the amino acid sequence of a *D. melanogaster* VRILLE isoform (Accession No. **AAF52237**; Adams et al., 2000), one transcript was identified as encoding a putative *C. finmarchicus* VRILLE (Calfi-VRI; Table 1). The protein deduced from this transcript is 467 amino acids long and appears to be a full-length protein. Alignment of the *Drosophila* and *Calanus* proteins is shown in Figure 15. As can be seen from this alignment, Calfi-VRI shares 22.6% identity and 42.4% similarity with its *Drosophila* counterpart.

Structural analysis of Calfi-VRI using the online program SMART identified one basic region leucine zipper domain within the protein (Fig. 15); the same motif was identified by this program in the *Drosophila* query (Fig. 15). The basic region leucine zippers of the *Calanus* and *Drosophila* proteins share much higher levels of amino acid conservation than do the two proteins when compared over their full lengths (*i.e.* 75.8% identity/93.9% similarity).

While the overall sequence conservation between Calfi-VRI and its *Drosophila* counterpart is low, BLASTing of the *Calanus* protein against the annotated sequences in FlyBase did identify an isoform of VRILLE, isoform C, (FlyBase No. **FBpp0078705**) as the most similar *D. melanogaster* protein to the query sequence. Likewise, reverse BLAST analysis against all non-redundant arthropod proteins in GenBank identified VRILLE from the butterfly *D. plexippus* (Accession No. **AAT86041**; Casselman and Reppert, unpublished submission) as the protein most similar to Calfi-VRI (Table 2). These data, in conjunction with those presented above, provide support for Calfi-VRI being a member of the VRILLE family.

3.1.3. Clock input proteins

3.1.3.1. CRYPTOCHROME 1 (CRY1): Using the amino acid sequence of *D. melanogaster* CRYPTOCHROME (Accession No. **AAC83828**; Emery et al., 1998), one *C. finmarchicus* transcript was identified as encoding a putative CRYPTOCHROME 1 (Calfi-CRY1; Table 1). The protein deduced from this transcript is 531 amino acids long and appears to be a full-length protein. Figure 16 shows the alignment of the *Drosophila* and *Calanus* proteins, and as can be seen here, Calfi-CRY1 shares 39.1% identity and 69.9%

similarity with the query sequence. No functional domains were identified in either Calfi-CRY1 or the *Drosophila* protein using SMART.

BLASTing of Calfi-CRY1 against the annotated proteins in FlyBase identified (6-4)-photolyase, isoform D, (FlyBase No. **FBpp0307794**) as the most similar *D. melanogaster* protein to the query sequence, while reverse BLAST analysis against all non-redundant arthropod proteins in GenBank identified DNA photolyase 1 from the water flea *D. pulex* (Accession No. **EFX85418**; Colbourne et al, 2011) as the sequence most similar to Calfi-CRY1 (Table 2). Because of the limited sequence conservation and the identification of photolyases as the most similar proteins in reverse BLAST analyses, it is inconclusive as to whether Calfi-CRY1 is an isoform of the CRYPTOCHROME 1 or a member of the photolyase family of proteins (and hence more appropriately named Calfi-photolyase).

3.2. Expressed sequence tag support for deduced protein sequences

The transcriptome used here to mine for *C. finmarchicus* circadian proteins was assembled *de novo* from a large number of 100-bp, paired-end reads. Given this, it is highly likely that some of the sequences produced by the Trinity assembly software may contain artifacts. If these errors are located within the coding region of a transcript, or introduce frame shifts upstream of the coding region, they may manifest themselves in errors in the protein sequences deduced from the transcriptome. To help vet the sequences of the proteins presented here, we used each deduced protein as a query for BLASTing against the extant *C. finmarchicus* ESTs curated in GenBank. While small in number relative to the sequences present in our Trinity assembly (~11,000 vs. ~200,000), the ESTs are all single-pass Sanger sequences (Lenz et al., 2012), and thus are not subject to assembly errors.

Figure 17 shows an example of the EST vetting of one of the Trinity-deduced circadian proteins. Specifically, the figure shows the alignment of Calfi-CRY2 and the four partial proteins that were derived from ESTs (Accession Nos. **EL966083**, **FK041536**, **FG985765** and **FK671009**) identified using the full-length sequence as the query (Table 3). As is evident from the alignment, the proteins deduced from **EL966083**, **FK041536**, **FG985765** are N-terminal partial sequences, while that derived from **FK671009** is an internal fragment (Fig. 17 and Table 3). The sequence deduced from **EL966083** is the longest of the N-terminal partial proteins, being 193 amino acids in length. This partial sequence is nearly identical to the corresponding portion of Calfi-CRY2, differing from it by, at most, 2 residues (there is one uncalled amino acid in the partial protein). The internal fragment deduced from **FK671009** (208 amino acids in length) is also nearly identical to its corresponding portion of Calfi-CRY2, differing from it at just one residue. Taken collectively, the EST-derived sequences encompass over 50% of the length of Calfi-CRY2 (401 of 794 amino acids), and vet all but one of the residues present in the regions of overlap. In addition to Calfi-CRY2, support for the sequence assignments of the Calfi-PER I, Calfi-TIM, Calfi-CKII α , Calfi-CWO, Calfi-PP1, Calfi-WBT I, Calfi-TWS, Calfi-SGG and Calfi-SLIMB proteins was obtained from the extant collection of *C. finmarchicus* ESTs (Table 3). Thus, nearly 50% of the *Calanus* circadian proteins deduced from the *de novo* transcriptome assembly were vetted, at least in part, using the extant EST sequences, a rather remarkable proportion of the proteins given the 20-fold difference in the number of sequences present in two data sets.

3.3. Identification of a *Calanus finmarchicus* EST encoding a putative partial CRYPTOCHROME 1

As stated in Section 3.1.2.4., no *C. finmarchicus* transcripts encoding homologs of *Drosophila* PAR DOMAIN PROTEIN 1e were identified within our *de novo* transcriptome assembly. Similarly, while a *Calanus* transcript was identified as being CRY1-like using a

Drosophila CRYPTOCHROME query, the protein deduced from it may well be a photolyase rather than a true CRY1 (see Section 3.1.3.1). Thus, two possibilities exist for PDP1 ϵ and CRY1 in *C. finmarchicus*: 1) the species does not possess a homolog of it, or 2), it is present in *Calanus*, but simply not among the sequences present in our transcriptome assembly. Given our ability to find good EST support for a number of our Trinity-derived proteins, we probed the extant *C. finmarchicus* ESTs for both proteins using the same *Drosophila* sequences used in Sections 3.1.2.4 and 3.1.3.1 to query our *de novo* transcriptome assembly. Here again, no transcript encoding a putative PDP1 ϵ was identified (data not shown). In contrast, one EST (Accession No. **GR410803**; E-value = 8e-61) was identified as encoding a putative CRY1. Translation of this EST showed it to encode a 298 amino acid N-terminal partial protein with 32.9% identity/62.6% similarity to the *Drosophila* query in their regions of overlap (Fig. 18). This level of conservation at the N-terminus is higher than that seen in a comparison of the same regions of Drome-CRY and the putative Calfi-CRY1 described in Section 3.1.3.1 (*i.e.* 29.5% identity/57.0% similarity; Fig. 16). Moreover, reverse BLAST analysis of the partial protein deduced from **GR410803** against the annotated proteins in FlyBase identified CRYPTOCHROME, isoform A, (FlyBase No. **FBpp0083150**; E-value = 8e-58) as the most similar *Drosophila* protein to the EST-derived query. Similarly, CRYPTOCHROME 1 from the silkworm *Antheraea pernyi* (Accession No. **AAK11644**; E-value = 1e-75; Yuan et al., 2007) was identified as the most similar protein to the query when reverse BLAST analysis against the non-redundant arthropod proteins curated in NCBI was run. Thus, there does appear to be a true CRYPTOCHROME 1 in *C. finmarchicus*, though the protein described in Section 3.1.3.1, Calfi-CRY1 from our Trinity assembly, may well not be a member of this protein group, but rather a protein more closely related to the (6-4)-photolyases.

4. Discussion

4.1. Identification of putative *Calanus finmarchicus* circadian proteins using a *de novo* transcriptome assembly

Many crustaceans are known to exhibit physiological and behavioral patterns that oscillate with a period of approximately 24-hours and are synchronized to the solar day. For example, circadian rhythms in locomotory behavior, molting, color change and reproductive control have all been described from members of this arthropod subphylum (*e.g.* Strauss and Dirksen, 2010). In contrast to the relative wealth of data on circadian phenomena in crustaceans, vanishingly little is known about the molecular underpinning of circadian rhythms in any member of this taxon, though their close phylogenetic relationship with the insects would suggest commonalities (*e.g.* Strauss and Dirksen, 2010). In fact, with the exception of a recent bioinformatics study conducted on the genome of the cladoceran *D. pulex* (Tilden et al., 2011), essentially nothing is known about the genes/proteins underlying circadian control in crustaceans (*e.g.* Strauss and Dirksen, 2010). Recently, we embarked upon a large-scale transcriptome sequencing project for the copepod crustacean *C. finmarchicus*, the biomass dominant zooplankton species for much of the North Atlantic Ocean (Christie et al., 2013a). Using Illumina sequencing and the software package Trinity, a transcriptome comprised of over 200,000 unique sequences was assembled *de novo* for *C. finmarchicus* (Christie et al., 2013a). Mining of this resource for peptidergic signaling transcripts, including those encoding both peptide precursor proteins and peptide receptors, revealed the assembly to be of high quality and to have good coverage, at least 60% based on the rate of discovery for peptidergic signaling proteins (Christie et al., 2013a). Given the number of sequences present in the assembly, the *C. finmarchicus* transcriptome produced by our lab represents one of, if not the, largest single-species resource for crustacean transcript discovery, including circadian protein-encoding ones, the subject of the study presented here.

Using the well-defined molecular clock of the fruit fly *D. melanogaster* as a template, we successfully identified *C. finmarchicus* transcripts encoding 16 of the 17 circadian protein families targeted, *i.e.* the core clock proteins CLOCK, CYCLE, PERIOD and TIMELESS, the clock-associated proteins CASEIN KINASE II α , CASEIN KINASE II β , CLOCKWORK ORANGE, DOUBLETIME, PROTEIN PHOSPHATASE 1, PROTEIN PHOSPHATASE 2A – MICROTUBULE STAR, PROTEIN PHOSPHATASE 2A – WIDERBORST, PROTEIN PHOSPHATASE 2A – TWINS, SHAGGY, SUPERNUMERARY LIMBS and VRILLE, and, possibly, the clock input pathway protein CRYPTOCHROME 1. In fact, only PAR DOMAIN PROTEIN 1 ϵ could not be found in the *Calanus* transcriptome. In addition, a *C. finmarchicus* transcript encoding the core clock protein CRYPTOCHROME 2 was identified using a sequence from the monarch butterfly *D. plexippus* as our query; *Drosophila* does not possess CRY2. Taken collectively, transcripts encoding 17 (16 if you exclude Calfi-CRY1) of the 18 circadian protein families queried for were successfully identified, for a discovery rate of 94%, 34% higher than our success rate for those encoding peptidergic signaling proteins. Moreover, 20 of the 21 proteins deduced from the identified transcripts appear to be full-length sequences, with approximately 50% of these strongly supported by EST evidence. Thus, the circadian protein discovery presented here not only provides the first description of the molecular components of a *C. finmarchicus* circadian system, but also continues to vet the quality and completeness of our *de novo* *Calanus* transcriptome and suggests that its coverage may be higher than the 60% reported previously (Christie et al., 2013a). Clearly, this transcriptome will be a valuable resource for transcript/protein discovery in *Calanus* for the foreseeable future, and lays a strong foundation for gene-based studies of the physiological ecology of this environmentally significant species.

4.2. Identification of the first set of “transcribed” crustacean circadian transcripts

In addition to being the first identification of the molecular components of a circadian clock in *C. finmarchicus*, the sequences identified here represent only the second “full” set of circadian transcripts/proteins reported for any crustacean, the first set being from the cladoceran *D. pulex* (Tilden et al., 2011). In fact, to the best of our knowledge, only two other crustacean circadian sequences have been published: an isoform of CLOCK from the freshwater prawn *Machrobrachium rosenbergii* (Yang et al., 2006), and a CRYPTOCHROME 2 from Antarctic krill *Euphausia superba* (Mazzotta et al., 2010).

A major difference between the *Calanus* data reported here and the *Daphnia* circadian dataset generated by Tilden et al. (2011) is that the former was obtained via transcriptome mining, and the latter via genome mining. Given that not all genes are transcribed, the *C. finmarchicus* sequences represent the first report of a “full” set of transcribed circadian proteins for any crustacean.

Moreover, the presence of both CRY1 and CRY2 in the *Calanus* transcriptome assembly suggests that *C. finmarchicus* possesses an ancestral-type circadian clock like that of the monarch butterfly *D. plexippus* (both CRYs present), rather than the more derived circadian system of *D. melanogaster*, where CRY2 has been lost. Given that both CRYs were also found within the *Daphnia* genome (Tilden et al., 2011) and a CRY2 has been cloned from the krill *E. superba* (Mazzotta et al., 2010), the emerging picture is that the circadian systems of crustaceans, at least basal ones, may all be of the ancestral-type, though clearly additional investigation is needed to confirm this hypothesis.

4.3. Conservation and variation in circadian protein sequences

As mentioned earlier, we successfully identified *C. finmarchicus* transcripts encoding homologs of 17 of the 18 circadian protein families that were targeted in our study.

Interestingly, the deduced *Calanus* proteins often differed considerably in amino acid composition from the *Drosophila*/*Danaus* proteins used to identify them, as well from their closest arthropod counterpart identified via reverse BLAST analyses. For example Calfi-PER II and Calfi-CWO, share only 24–25% overall amino acid identity with the arthropod proteins identified as most similar to them (in these cases proteins from the cricket *A. asahinai* and the ant *C. floridanus*, respectively).

Interestingly, in the functional domains predicted by SMART, the degree of amino acid conservation often rose considerably between protein pair (e.g. the helix-loop-helix and orange domains of Calfi-CWO share 80.7%/96.5% and 31.9%/70.2% amino acid identity/similarity with those of the *Drosophila* protein used to identify it, while the two proteins compared in their entirety share only 24.8% identity/51.3% similarity), and for many of our identifications, this sort of vetting was crucial for convincing us that we truly had deduced a *C. finmarchicus* member of a target family. Similar sequence variation was noted by Tilden et al. (2011) in identifying and characterizing circadian proteins from the genome of *D. pulex*, and here too reverse BLASTing was crucial for confirming the identity of the deduced sequences. Thus, the multipronged strategy used here for identifying and vetting the circadian genes/proteins of *Calanus* will likely serve as a useful model for future bioinformatics-based elucidations of circadian systems from other crustacean species.

4.4. Biological rhythmicity in *Calanus finmarchicus*

Given our identification of an essentially full complement of circadian transcripts/proteins from *C. finmarchicus*, it seems highly likely that these molecules form the basis of a molecular clock in this species. Where this clock, or clocks, is located in the organism is an open question, though homology to insects would suggest at least one neuronal clock is present in the brain (Allada and Chung, 2010; Tomioka and Matsumoto, 2010). What physiological/behavioral patterns are controlled by the circadian system of *Calanus* is also an open question.

Copepods are known to exhibit circadian rhythmicity in feeding, vertical migration, reproduction, swimming activity and regeneration of gut epithelium (Mauchline, 1998). However, the relative importance of exogenous and endogenous control of these behaviors is still poorly understood (for review see: Naylor, 2006; Cohen and Forward, 2009; Marcus and Sheef, 2009). Copepod circadian patterns can be quite variable along temporal and spatial gradients within a single species, as well as among species. Experimental and field studies have linked diel vertical migration and feeding patterns to exogenous factors, such as light and predation risk (Huntley and Brooks, 1982; Daro, 1985; Calbet et al., 1999; Cieri and Stearns, 1999; Cohen and Forward, 2005a; Vopel and Thistle, 2011). However, endogenous rhythms in feeding, reproduction, bioluminescent activity and swimming have also been demonstrated in copepods under controlled experimental conditions (Enright and Hamner, 1967; Umminger, 1968; Marcus, 1985; Bowlby and Case, 1989; Calbet et al., 1999; Seuront 2011). These contrasting results may be a reflection of behavioral plasticity (Cohen and Forward, 2009). Cohen and Forward (2005b) found that diel vertical migration behavior showed endogenous rhythms, but was also easily entrained by changes in absolute irradiance under experimental conditions. Estuarine copepods, such as *Eurytemora affinis* have also been shown to exhibit endogenous circatidal rhythms in swimming activity (Hough and Naylor, 1992).

Daily and seasonal rhythms documented for *C. finmarchicus* include diel vertical migration and feeding activity, ontogenetic vertical migration and seasonal diapause (e.g. Hirche, 1996; Hays, 2003; Dale and Kaartvedt, 2000; Baumgartner et al., 2011; Rabindranath et al., 2011). It is an open question as to whether some or all of these behavioral patterns are under endogenous control and whether they are entrained by light, temperature or other

environmental factors (e.g. Marcus and Sheef, 2009). The predicted clock proteins presented here will open opportunities to experimentally address questions on which behaviors are controlled by the circadian clock, as well as its involvement in other seasonal rhythms. Molecular and cellular techniques, such as *in situ* hybridization, immunohistochemistry and RNAseq, can easily be applied to small organisms such as *C. finmarchicus* to test hypotheses on the relative roles of phenotypic plasticity, photoperiodism and exogenous factors in driving behavior in this species.

5. Conclusions

Via Illumina sequencing and *de novo* assembly, a transcriptome consisting of over 200,000 sequences was produced for the copepod *C. finmarchicus* (Christie et al., 2013a). This resource has allowed for the mining of transcripts that encode proteins putatively involved in the establishment of a circadian clock in this ecologically critical species. Given the continuing decrease in cost of high throughput sequencing, and the plethora of free, publicly accessible software for assembling and probing these data, the production and mining of *de novo* transcriptome assemblies now provide a new strategy for identifying the molecular component of biological clocks in other crustaceans as well. Clearly circadian systems play important roles in crustaceans, and understanding their generation, maintenance and modulation will provide critical insights for understanding physiological and behavioral control in these animals. As circadian clocks have been constructed in multiple ways in the Insecta (e.g. Yuan et al., 2007), it is possible that crustaceans too show variation in the genes and proteins underlying their circadian rhythms. As more data are generated for members of this arthropod subphylum, it will be interesting to see how great this variation may be, as well as what common threads exist across the circadian systems of this and other members of the Arthropoda. Due to the number and diversity of species that comprise this phylum, the elucidation of common and unique features of circadian clocks in these animals is likely to shed important light on the evolution biological clocks in a general sense, including whether or not clocks operating on different timing cycles (e.g. circadian and seasonal clocks) use the same or distinct molecular machinery to affect physiological/behavioral output.

Acknowledgments

We would like to thank Vittoria Roncalli (University of Hawaii at Manoa), R. Patrick Hassett (Ohio University), Roy McMorran (Mount Desert Island Biological Laboratory [MDIBL]), Christine Smith (MDIBL), Benjamin King (MDIBL) and Daniel Hartline (University of Hawaii at Manoa) for sharing resources and helpful discussions concerning this study. Richard LeDuc, Le-Shin Wu, Carrie Ganote, and Thomas Doak (National Center for Genome Analysis Support, Indiana University) are thanked for their assistance in producing the *C. finmarchicus* transcriptome used in this study. The authors acknowledge financial support from The National Science Foundation (NSF OCE 1040597), the National Institutes of Health National Institute of General Medical Sciences (NTH 2T34GM007684-32 to Patricia A. Couvillon and Petra H. Lenz), and the Cades Foundation of Honolulu Hawaii.

References

- Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW, Hoskins RA, Galle RF, George RA, Lewis SE, Richards S, Ashburner M, Henderson SN, Sutton GG, Wortman JR, Yandell MD, Zhang Q, Chen LX, Brandon RC, Rogers YH, Blazej RG, Champe M, Pfeiffer BD, Wan KH, Doyle C, Baxter EG, Helt G, Nelson CR, Gabor GL, Abril JF, Agbayani A, An HJ, Andrews-Pfannkoch C, Baldwin D, Ballew RM, Basu A, Baxendale J, Bayraktaroglu L, Beasley EM, Beeson KY, Benos PV, Berman BP, Bhandari D, Bolshakov S, Borkova D, Botchan MR, Bouck J, Brokstein P, Brottier P, Burtis KC, Busam DA, Butler H, Cadieu E, Center A, Chandra I, Cherry JM, Cawley S, Dahlke C, Davenport LB, Davies P, de Pablos B, Delcher A, Deng Z, Mays AD, Dew I, Dietz SM, Dodson K, Doup LE, Downes M, Dugan-Rocha S, Dunkov BC, Dunn P, Durbin KJ, Evangelista CC, Ferraz C, Ferriera S, Fleischmann W, Fosler C, Gabrielian AE, Garg NS, Gelbart WM, Glasser K, Glodek A, Gong F, Gorrell JH, Gu Z, Guan P, Harris M, Harris NL, Harvey D, Heiman TJ, Hernandez JR, Houck J, Hostin D, Houston KA, Howland TJ,

Wei MH, Ibegwam C, Jalali M, Kalush F, Karpen GH, Ke Z, Kennison JA, Ketchum KA, Kimmel BE, Kodira CD, Kraft C, Kravitz S, Kulp D, Lai Z, Lasko P, Lei Y, Levitsky AA, Li J, Li Z, Liang Y, Lin X, Liu X, Mattei B, McIntosh TC, McLeod MP, McPherson D, Merkulov G, Milshina NV, Mobarry C, Morris J, Moshrefi A, Mount SM, Moy M, Murphy B, Murphy L, Muzny DM, Nelson DL, Nelson DR, Nelson KA, Nixon K, Nusskern DR, Pacleb JM, Palazzolo M, Pittman GS, Pan S, Pollard J, Puri V, Reese MG, Reinert K, Remington K, Saunders RD, Scheeler F, Shen H, Shue BC, Sidén-Kiamos I, Simpson M, Skupski MP, Smith T, Spier E, Spradling AC, Stapleton M, Strong R, Sun E, Svirskas R, Tector C, Turner R, Venter E, Wang AH, Wang X, Wang ZY, Wassarman DA, Weinstock GM, Weissenbach J, Williams SM, Woodage T, Worley KC, Wu D, Yang S, Yao QA, Ye J, Yeh RF, Zaveri JS, Zhan M, Zhang G, Zhao Q, Zheng L, Zheng XH, Zhong FN, Zhong W, Zhou X, Zhu S, Zhu X, Smith HO, Gibbs RA, Myers EW, Rubin GM, Venter JC. The genome sequence of *Drosophila melanogaster*. Science. 2000; 287:2185–2195. [PubMed: 10731132]

- Allada R, Chung BY. Circadian organization of behavior and physiology in *Drosophila*. Annu Rev Physiol. 2010; 72:605–624. [PubMed: 20148690]
- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997; 25:3389–3402. [PubMed: 9254694]
- Bae K, Lee C, Sidote D, Chuang KY, Edery I. Circadian regulation of a *Drosophila* homolog of the mammalian Clock gene: PER and TIM function as positive regulators. Mol Cell Biol. 1998; 18:6142–6151. [PubMed: 9742131]
- Baumgartner MF, Lysiak NSJ, Schuman C, Urban-Rich J, Wenzel FW. Diel vertical migration behavior of *Calanus finmarchicus* and its influence on right and sei whale occurrence. Mar Ecol Prog Ser. 2011; 423:167–184.
- Bonasio R, Zhang G, Ye C, Mutti NS, Fang X, Qin N, Donahue G, Yang P, Li Q, Li C, Zhang P, Huang Z, Berger SL, Reinberg D, Wang J, Liebig J. Genomic comparison of the ants *Camponotus floridanus* and *Harpegnathos saltator*. Science. 2010; 329:1068–1071. [PubMed: 20798317]
- Bowlby MR, Case JF. Endogenous bioluminescent characteristics of a mesopelagic copepod. Am Zool. 1989; 29:A37.
- Bünning E. Die endogene Tagesrhythmik als Grundlage der Photoperiodischen Reaktion. Ber Deutschen Bot Gesellsch. 1936; 54:590–607.
- Calbet A, Saiz E, Irigoien X, Alcaraz M, Trepas I. Food availability and diel feeding rhythms in the marine copepods *Acartia grani* and *Centropages typicus*. J Plankton Res. 1999; 21:1009–1015.
- Christie AE, Roncalli V, Lona PB, McCool MD, King BL, Bucklin A, Hartline DK, Lenz PH. *In silico* characterization of the insect diapause-associated protein couch potato (CPO) in *Calanus finmarchicus* (Crustacea: Copepoda). Comp Biochem Physiol D. 2013b; 8:45–57.
- Christie AE, Roncalli V, Wu L, Ganote CL, Doak T, Lenz PH. Peptidergic signaling in *Calanus finmarchicus* (Crustacea, Copepoda): *in silico* identification of putative peptide hormones and their receptors using a *de novo* assembled transcriptome. Gen Comp Endocrinol. 2013a In press.
- Cieri MD, Stearns DE. Reduction of grazing activity of two estuarine copepods in response to the exudate of a visual predator. Mar Ecol Prog Ser. 1999; 177:157–163.
- Cohen JH, Forward RB Jr. Photobehavior as an inducible defense in the marine copepod *Calanopia americana*. Limnol Oceanogr. 2005a; 50:1269–1277.
- Cohen JH, Forward RB Jr. Diel vertical migration of the marine copepod *Calanopia americana*. II. Proximate role of exogenous light cues and endogenous rhythms. Mar Biol. 2005b; 147:399–410.
- Cohen JH, Forward RB Jr. Zooplankton diel vertical migration - a review of proximate control. Oceanogr Mar Biol Annu Rev. 2009; 47:77–109.
- Colbourne JK, Pfreder ME, Gilbert D, Thomas WK, Tucker A, Oakley TH, Tokishita S, Aerts A, Arnold GJ, Basu MK, Bauer DJ, Cáceres CE, Carmel L, Casola C, Choi JH, Dettler JC, Dong Q, Dusheyko S, Eads BD, Fröhlich T, Geiler-Samerotte KA, Gerlach D, Hatcher P, Jogdeo S, Krijgsveld J, Kriventseva EV, Kültz D, Laforsch C, Lindquist E, Lopez J, Manak JR, Muller J, Pangilinan J, Patwardhan RP, Pitluck S, Pritham EJ, Rechtsteiner A, Rho M, Rogozin IB, Sakarya O, Salamov A, Schaack S, Shapiro H, Shiga Y, Skallitzky C, Smith Z, Souvorov A, Sung W, Tang Z, Tsuchiya D, Tu H, Vos H, Wang M, Wolf YI, Yamagata H, Yamada T, Ye Y, Shaw JR, Andrews J, Crease TJ, Tang H, Lucas SM, Robertson HM, Bork P, Koonin EV, Zdobnov EM,

- Grigoriev IV, Lynch M, Boore JL. The ecoresponsive genome of *Daphnia pulex*. *Science*. 2011; 331:555–561. [PubMed: 21292972]
- Dale T, Kaartvedt S. Diel patterns in stage-specific vertical migration of *Calanus finmarchicus* in habitats with midnight sun. *ICES J Mar Sci*. 2000; 57:1800–1818.
- Daro MH. Feeding rhythms and vertical distribution of marine copepods. *Bull Mar Sci*. 1985; 37:487–497.
- de laIglesia HO, Hsu YW. Biological clocks and rhythms in intertidal crustaceans. *Front Biosci (Elite Ed)*. 2010; 2:1394–1404. [PubMed: 20515812]
- Dombrádi V, Axton JM, Brewis ND, da Cruz e Silva EF, Alphey L, Cohen PT. *Drosophila* contains three genes that encode distinct isoforms of protein phosphatase 1. *Eur J Biochem*. 1990; 194:739–745. [PubMed: 2176604]
- Emery P, So WV, Kaneko M, Hall JC, Rosbash M. CRY, a *Drosophila* clock and lightregulated cryptochrome, is a major contributor to circadian rhythm resetting and photosensitivity. *Cell*. 1998; 95:669–679. [PubMed: 9845369]
- Enright TJ, Hamner WM. Vertical diurnal migration and endogenous rhythmicity. *Science*. 1967; 157:937–941. [PubMed: 17792830]
- Farca Luna AJ, Heinrich R, Reischig T. The circadian biology of the marbled crayfish. *Front Biosci (Elite Ed)*. 2010; 2:1414–1431. [PubMed: 20515814]
- Giribet G, Edgecombe GD. Reevaluating the arthropod tree of life. *Annu Rev Entomol*. 2012; 57:167–186. [PubMed: 21910637]
- Hays GC. A review of the adaptive significance and ecosystem consequences of zooplankton diel vertical migrations. *Hydrobiologia*. 2003; 503:163–170.
- Hirche HJ. Diapause in the marine copepod *Calanus finmarchicus* – a review. *Ophelia*. 1996; 44:129–143.
- Hough AR, Naylor E. Field studies on retention of the planktonic copepod *Eurytemora affinis* in a mixed estuary. *Mar Ecol Prog Ser*. 1991; 76:115–122.
- Huntley M, Brooks ER. Effects of age and food availability on diel vertical migration of *Calanus pacificus*. *Mar Biol*. 1982; 71:23–31.
- Kamae Y, Tanaka F, Tomioka K. Molecular cloning and functional analysis of the clock genes, Clock and cycle, in the firebrat *Thermobia domestica*. *J Insect Physiol*. 2010; 56:1291–1299. [PubMed: 20416313]
- Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res*. 2002; 30:3059–3066. [PubMed: 12136088]
- Katoh K, Toh H. Recent developments in the MAFFT multiple sequence alignment program. *Brief Bioinform*. 2008; 9:286–298. [PubMed: 18372315]
- Lenz PH, Unal E, Hassett RP, Smith CM, Bucklin A, Christie AE, Towle DW. Functional genomics resources for the North Atlantic copepod, *Calanus finmarchicus*: EST database and physiological microarray. *Comp Biochem Physiol D*. 2012; 7:110–123.
- Letunic I, Doerks T, Bork P. SMART 6: Recent updates and new developments. *Nucleic Acids Res*. 2009; 37:D229–D232. [PubMed: 18978020]
- Lin SC, Lin MH, Horváth P, Reddy KL, Storti RV. PDP1, a novel *Drosophila* PAR domain bZIP transcription factor expressed in developing mesoderm, endoderm and ectoderm, is a transcriptional regulator of somatic muscle genes. *Development*. 1997; 124:4685–4696. [PubMed: 9409684]
- Marcus MH. Endogenous control of spawning in a marine copepod. *J Exp Mar Biol Ecol*. 1985; 91:263–269.
- Marcus, MH.; Sheef, LP. Photoperiodism in copepods. In: Nelson, RJ.; Denlinger, DL.; Somers, DE., editors. *Photoperiodism: The Biological Calendar*. Oxford University Press; Oxford: 2009. p. 193–217.
- Mauchline, J. *Advances in Marine Biology – The Biology of Calanoid Copepods*. Academic Press; San Diego: 1998.

- Mazzotta GM, De Pittà C, Benna C, Tosatto SC, Lanfranchi G, Bertolucci C, Costa R. A cry from the krill. *Chronobiol Int.* 2010; 27:425–445. [PubMed: 20524794]
- McCoole MD, Atkinson NJ, Graham DI, Grasser EB, Joselow AL, McCall NM, Welker AM, Wilsterman EJ Jr, Baer KN, Tilden AR, Christie AE. Genomic analyses of aminergic signaling systems (dopamine, octopamine and serotonin) in *Daphnia pulex*. *Comp Biochem Physiol D.* 2012a; 7:35–58.
- McCoole MD, Baer KN, Christie AE. Histaminergic signaling in the central nervous system of *Daphnia* and a role for it in the control of phototactic behavior. *J Exp Biol.* 2011; 214:1773–1782. [PubMed: 21525325]
- McCoole MD, D'Andrea BT, Baer KN, Christie AE. Genomic analyses of gas (nitric oxide and carbon monoxide) and small molecule transmitter (acetylcholine, glutamate and GABA) signaling systems in *Daphnia pulex*. *Comp Biochem Physiol D.* 2012b; 7:124–160.
- Myers MP, Wager-Smith K, Wesley CS, Young MW, Sehgal A. Positional cloning and sequence analysis of the *Drosophila* clock gene, timeless. *Science.* 1995; 270:805–808. [PubMed: 7481771]
- Naylor E. Orientation and navigation in coastal and estuarine zooplankton. *Mar Freshw Behav Physiol.* 2006; 39:13–24.
- Nelson, R.J.; Denlinger, D.L.; Somers, D.E. *Photoperiodism: The Biological Calendar.* Oxford University Press; 2010. p. 581
- Rabindranath A, Daase M, Falk-Petersen S, Wold A, Wallace MI, Berge J, Brierley AS. Seasonal and diel vertical migration of zooplankton in the High Arctic during the autumn midnight sun of 2008. *Mar Biodiv.* 2011; 41:365–382.
- Saunders DS. Evidence for 'dawn' and 'dusk' oscillators in the *Nasonia* photoperiodic clock. *J Insect Physiol.* 1974; 20:77–88.
- Saunders DS. Controversial aspects of photoperiodism in insects and mites. *J Insect Physiol.* 2010; 56:1491–1502. [PubMed: 20471388]
- Saunders DS. Insect photoperiodism: Measuring the night. *J Insect Physiol.* 2013; 59:1–10. [PubMed: 23183333]
- Saunders DS, Bertossa RC. Deciphering time measurement: the role of circadian 'clock' genes and formal experimentation in insect photoperiodism. *J Insect Physiol.* 2011; 57:557–566. [PubMed: 21295039]
- Schultz J, Milpetz F, Bork P, Ponting CP. SMART, a simple modular architecture research tool: identification of signaling domains. *Proc Natl Acad Sci USA.* 1998; 26:5857–5864. [PubMed: 9600884]
- Seuront L. Behavioral fractality in marine copepods endogenous rhythms versus exogenous stressors. *Physica A.* 2011; 390:250–256.
- Strauss J, Dirksen H. Circadian clocks in crustaceans: identified neuronal and cellular systems. *Front Biosci.* 2010; 15:1040–1074.
- Tilden AR, McCoole MD, Harmon SM, Baer KN, Christie AE. Genomic identification of a putative circadian system in the cladoceran crustacean *Daphnia pulex*. *Comp Biochem Physiol D.* 2011; 6:282–309.
- Tomioka K, Matsumoto A. A comparative view of insect circadian clock systems. *Cell Mol Life Sci.* 2010; 67:1397–1406. [PubMed: 20035363]
- Umminger BL. Polarotaxis in copepods. I. An endogenous rhythm in polarotaxis in *Cyclops vernalis* and its relation to vertical migration. *Biol Bull.* 1968; 135:239–251.
- Uryu O, Tomioka K. Circadian oscillations outside the optic lobe in the cricket *Gryllus bimaculatus*. *J Insect Physiol.* 2010; 56:1284–1290. [PubMed: 20416318]
- Vopel K, Thistle D. Cues, not an endogenous rhythm, control the dusk peak in water-column entry by benthic copepods. *Estuaries Coasts.* 2011; 34:1194–1204.
- Werckenthin A, Derst C, Stengl M. Sequence and expression of per, tim1, and cry2 genes in the Madeira cockroach *Rhyparobia maderae*. *J Biol Rhythms.* 2012; 27:453–466. [PubMed: 23223371]
- Yang JS, Dai ZM, Yang F, Yang WJ. Molecular cloning of Clock cDNA from the prawn, *Macrobrachium rosenbergii*. *Brain Res.* 2006; 1067:13–24. [PubMed: 16271708]

- Yuan Q, Metterville D, Briscoe AD, Reppert SM. Insect cryptochromes: gene duplication and loss define diverse ways to construct insect circadian clocks. *Mol Biol Evol.* 2007; 24:948–955. [PubMed: 17244599]
- Zhu H, Sauman I, Yuan Q, Casselman A, Emery-Le M, Emery P, Reppert SM. Cryptochromes define a novel circadian clock mechanism in monarch butterflies that may underlie sun compass navigation. *PLoS Biol.* 2008; 6:e4. [PubMed: 18184036]
- Zhu H, Yuan Q, Briscoe AD, Froy O, Casselman A, Reppert SM. The two CRYs of the butterfly. *Curr Biol.* 2005; 15:R953–R954. [PubMed: 16332522]

***Drosophila melanogaster* vs. *Calanus finmarchicus* CLOCK**

Drome-CLK M-----DDESDDKDDTKSFLCRKSRNLSEKKRRDQFNSLVDDLSALI
 Calfi-CLK MSVSRGGRNSISMSVDFGDEGDGDEKDETK---RKSRLNLSEKKRRDQFNMVLTELGTMV
 * :.:*:*:*:*:* ***** * :*.:

 Drome-CLK STSSRKMDKSTVLKSTIAFLKNHNEATDRSKVFEIQDQWKPFLSNDEYTHLMLESIDGDF
 Calfi-CLK SPNNRKMDKTTVLKTTIAFLKQHNEMSIRSQAEEIQTQWKPFLSNDEYTHLMLEALDGF
 *...*****:*:*:*:*:* : *.: * * * * * :*:*:*:*:*

 Drome-CLK MMVFSSMSGISFYASESITSQLGYLPQDLNMTIYDLAYEMDHEALLNIFNMPTFVIEPRQ
 Calfi-CLK IIVLGTGHLIYTESMASLLGYMPTSLHNTLYEIVSDNDKIPLYNTLNMSASV-----
 :*.: : * *:*:*:*:* * *:* * : * : * * : :.*

 Drome-CLK TDISSNQITFYTHLRGMEKVDANAYELVKFVGFRNDTNTSTGSSSESVN-----
 Calfi-CLK -DTCGQQQLKLLIHIRRSCVSSKDS-LLELVLSGYFRKWNPAETPDYFSDDDDKISL
 * .:.*.: : *:* . : * : * : : * : * : * : * : * : * : * :

 Drome-CLK --GSNGQPAVLPRIFQONPNAEVDKLVFVGTGRVQNPQIREMSIIDPTSNEFTSKHSM
 Calfi-CLK KSGISVQSSQVQRF--RLPQAD--KTVEFATGRLATPQLRELFPVHVQSQSEFTSRHSL
 * . *.: : * : : * : * * * * * : * : * : * : * : * : * :

 Drome-CLK EWKFLFLDHRAPPIIGYMPFEVLGTSGYDYHFDLDSIVACHEELRQTGEGKSCYYRFL
 Calfi-CLK EWKFLFVDHRAPPIIGYLTPELLGTSGYDYHVDLEQVAVCHESIMQTGECTSCYYRFL
 *****:*

 Drome-CLK TKGQQWIWLQTDYYSYHQFNSKPDYVVCVTHKVVSYAEVLKDSRKEGQKSGNSNITNG
 Calfi-CLK TKGQQWIWLQTRYIITYHQWNSKPEFIVATHKVVNYKEVLNQATNPDKTDELNDQLNDSI
 ***** * : * : * : * : * : * : * : * : * : * : * : * : * : * : * :

 Drome-CLK SSKVIASGTSSKSASATTLRDFELSSQNLDSTLLGNSLASLGTETAATSPAVDSSPMW
 Calfi-CLK SEK--HSKTHSFRSGSPTWSSK-----SSLCGSDRTGFASSSRMSHNSNDNPMW
 * * * * * : * : * : * : * : * : * : * : * : * : * : * : * :

 Drome-CLK SASAVQPSGSCQINPLKTSRPASSYGNISSTGISPKAKRKCYFYNNRGNDSDS---TSM
 Calfi-CLK SVGQPAYRQSRLSHNRKHDRQTDSMAGMSE-----DMSCEEQNTDQTDAYSVTSETKF
 * . * : * * : * * : * . * * . * : * * :

 Drome-CLK STDVTSRQSMTHVSSQSQRSHHREHREHNNHNSHHMQQQQQHONQQQQHQQHQ
 Calfi-CLK SGPAVGDPPEPMSNDKPYLDQRQGEQHAEE-----
 * : * . : * . . . * * : * *

 Drome-CLK LQQQLQHTVGTPKMVPLPIASTQIMAGNACQFPQPAYPLASPQLVAPTLEPPQYLTAI
 Calfi-CLK -----KFL-----AV
 . * * *

 Drome-CLK PMQFVIAPFPVAPVLSPLPVQSQTDMPLDVTVMPTQSQLQDQLQRKHDELQKLILQQQN
 Calfi-CLK P-----
 *

 Drome-CLK ELRIVSEQLLSRYTYLQPMMSGFAPGNMTAAAVGNLGASGQRGLNFTGSNAVQPQFNQ
 Calfi-CLK -----

 Drome-CLK YGFALNSEQMLNQDQDQMMQQQQNLHTQHQNHLQQQHSHSQLQHTQQHQHQHQHQHQ
 Calfi-CLK -----

 Drome-CLK QQQQQQQQQQQQQQQQQQQQQQQQQQQQLQQLQQNDILLREDIDDIDAFLNLSPLHSLGSQST
 Calfi-CLK -----

 Drome-CLK INPFNSSNNNNQSYNGGSNLNNGNQNNNNRNSNPPQNNNEDSLLSCMQMATESSPSINF
 Calfi-CLK -----

 Drome-CLK HMGISDDGSETQSEDNKMHTSGSNLVQQQQQQQQQQILQQHQQQSNSFFSSNPFLNSQ
 Calfi-CLK -----

 Drome-CLK NQNQNQLPNDLEILPYQMSQEQSQNLFSPTAPGSSQ
 Calfi-CLK -----Q
 *

Figure 1.

Putative *Calanus finmarchicus* CLOCK (CLK) protein. Alignment of *Drosophila melanogaster* CLK (Drome-CLK; Accession No. AAC62234) with the partial *C. finmarchicus* CLK (Calfi-CLK) deduced from the Trinity *de novo* transcriptome assembly. In the line immediately below each sequence grouping, “*” indicates amino acids that are identical in the two proteins, while “.” and “:” denote amino acids that are similar in structure between the two sequences. In this figure, helix-loop-helix, PAS, PAC, and coiled coil domains identified by SMART analyses are highlighted in yellow, light green, dark green and gray, respectively.

***Danaus plexippus* vs. *Calanus finmarchicus* CRYPTOCHROME 2**

```

Danpl-CRY-2 MSVAETLPLRARSPTAQKSSQPAQVPEKEHTVHWFRKGLRLHDNPALREGLVDATTFRCV
Calfi-CRY-2 MT-----RKTSSRKLTVHWFRKGLRLSDNPALLRGIRRCETFCI
*:          .:. * ***** * * * * *: . * * *:

Danpl-CRY-2 FIIDPWFASSNVGINKWRFLQLQLEDLTKNLRLKLSRLFVVRGQPADALPKLFREWGTT
Calfi-CRY-2 FILDWFAGSSNVGTNKWRFLQLQLEDLTKSLRLKLSRLFVIRGQPADVLPNLFKQWGT
*: * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

Danpl-CRY-2 ALTFEEDPEPYGRVRDHNIMTKCREVGIQVTSRVSHSTLYKLDDIIEKNGGKAPLTYHQFQ
Calfi-CRY-2 YFTFEEDPEPFGRVRDQNIIVAMCKEMNIAVTYEHSTLYNLDKIERNNNKAPLTYKQFQ
*: * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

Danpl-CRY-2 ALIASMPPPPSAEPTISLETINRAVTPISDNHDERFGVPTLEELGFDTEGLKPPIIWIGGE
Calfi-CRY-2 RVIEGMEPPPKPVPTLTQLMLGSAVTPLTDDHDEKYGVPTLEELGFDCENLKAPVWTGGE
*: * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

Danpl-CRY-2 NEALLRLERHLERKAWVASFGRPKMTPESLLSSQTGLSPYLRFGLSTRLFYQSLSELYK
Calfi-CRY-2 TEALSRLERHLERKAWVASFGRPKMTPQSLLASQTGLSPYLRFGLSTRLFYHALSDLYR
*: * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

Danpl-CRY-2 RIKQERPPSLHGGIILWREFFYCAATRNPNFDRMEGNPICVQIPEWKNQEALKKWANGQT
Calfi-CRY-2 KIKKCDPPLSLHGGIILWREFFYCAATKNPNFDRKMSGNPICVQIIPWDNNNEALAKWANAH
*: * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

Danpl-CRY-2 GFPWIDAIMIQLRNDGWIHHLARHAVACFLTRGDLWISWEEGMKVFDELLLDADWSVNAG
Calfi-CRY-2 GFPWIDAIMTQLREEGWIHHLARHAVACFLTRGDLWVSWEEGLKVFDELLLDADWSVNAG
*: * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

Danpl-CRY-2 MWWLSCSSFFQFFHCYCPVRFGRKTDPNNGDFIRKYIPVLKNMPTRYIHEPWVCPEEIQ
Calfi-CRY-2 TWLWLSLSCSSFFQFFHCYCPVRFGRKADANGDFIRKYLPAKNFPTRYIHEPWTAPEAVQ
*: * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

Danpl-CRY-2 KSIRCIIGKDYPMPIVDHTKASEINLERIKQVYAQLAKFKPQGA-----LIPQMLQR
Calfi-CRY-2 KSAKCIIGQDYKPMCNHAYVSKLNMERMKQVYAQLAHYRRTGQTGPQIPQDLSQMKKL
*: * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

Danpl-CRY-2 PNV-----LQSSPSPTSIIANINQSNYLSQSSDVPTPTNQTQFKEDAVFLKPTVNN
Calfi-CRY-2 PKIPSPGGGLMGPPNPNLDSKKGRLHY-DKLASDQTDPNVQMQYLSNY-----QIQN
*: * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

Danpl-CRY-2 IKSNDVKQQQFQVIVQEDKHSENQRHVSNGKYIVNEINKNINDIPVKQNNYDFKALT
Calfi-CRY-2 LR-----QQGMSPALVQQPSRSQGQPPNMTTLQGYTQANR-IVPHQRENILLMDSPGM
*: * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

Danpl-CRY-2 NLNKF---SNEPLTFLNQTPNKNESFGQDVNNVIDVYSTSKPKFYFTDNGVITHNENAQT
Calfi-CRY-2 PQGQFIMTSRDQPMFGASTFKSEPSIPSSIIN-----PTTTPSIAGSSSGY---SSFSN
*: * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

Danpl-CRY-2 FKRDYSYD-NYNKESTGSNRVGEVHSNNPQTDKISS-----
Calfi-CRY-2 FNSCSLSDIKYASSSLPVISGDILLPGPISSTADSIISGATTIVEGCSTISNLSIMSQP
*: * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

Danpl-CRY-2 -----EKKN
Calfi-CRY-2 GSVEHIPSSLSPPVNMGTGSASQENLSSNTPIIGLDNYDAALAEGFFDHSAQASQN
*:

```

Figure 2.

Putative *Calanus finmarchicus* CRYPTOCHROME 2 (CRY2) protein. Alignment of *Danaus plexippus* CRY2 (Danpl-CRY2; Accession No. **ABA62409**) with the *C. finmarchicus* CRY2 (Calfi-CRY2) deduced from the Trinity *de novo* transcriptome assembly. In the line immediately below each sequence grouping, “*” indicates amino acids that are identical in the two proteins, while “.” and “:” denote amino acids that are similar in structure between the two sequences. No functional domains were identified by SMART analyses in either protein.

***Drosophila melanogaster* vs. *Calanus finmarchicus* CYCLE**

```

Drome-CYC      MEVQ-----EFCENMEEIEDENYDEEKSAR--TSDENRK--
Calfi-CYC      MEMYPQPDKMMTNTQPTENRRRRREDTIEDQEQMSRRRRRTIDDOEQMRIMPPDEQKKYL
                **:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*
Drome-CYC      -QNHSEIEKRRRDKMNTYINELSSMIPMCFAMQKRLDKLTVLRMAVQHRLRGIRGSGSLHP
Calfi-CYC      RENHSQIEKRRRDKMNTYITELCTMIPTCKAMSRKTDKLTVLRFAVQHLKTIR--GSLDS
                :***:*****:***:***:***:***:***:***:***:***:***:***:***:***:***:
Drome-CYC      FNGSDYRPSFLSDQELKMIILQASEGFLFVVGCDRGRILYVSDSVSSVLNSTQADLLGQS
Calfi-CYC      YTEGNYKPPMLTDKELKQLIIPCADGFMFVVD SARGRILFVSESVSDTLNFSQDLIGQS
                :. :*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:
Drome-CYC      WFDVLHPKDIGKVKEQLSSLEQCPRERLIDAKTMLPVK-TDVPQSLCRLCPGARRSFFCR
Calfi-CYC      LFDILHPKDIPKVKEQLSNSDSTPRERLIDSKTMLPLKGGDVPQSI SRLHPGARRVFFCR
                **:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*
Drome-CYC      MKLRTASNQIKEESD-----TSSSSRSSTKRKSLRTTGHKYRVIQCTGYLKS WT
Calfi-CYC      MKCK-QTLNPVKQEDDPNSSPMLNSTYSSSSSTAGKKKKSSLEKKYLSIQCTGYLKS WP
                ** : : * :*:*: * * * * : : * : * : : * * * * * * * * *
Drome-CYC      PIK---DED-QDADSDEQTTNLSCLVAIGRIPPVNRNSTVPASLDNHPNIRHVLFI SRHS
Calfi-CYC      LTKVGLDQDFSDPDNQGDSCLVAVGRIQPSFKTTIEDCLDKGQTTAPGVEFTSRHG
                * *:* *.*.*: : :*****:* *..*: : . . . . * * * *
Drome-CYC      GEGKFLFIDQRATLVIGFLPQEILGTSFYEFHNE DIAALMESHKVMVQVPEKVTQVYR
Calfi-CYC      TDGKFSYVDPRVTLLGYLPQELNGTSLYEHIQFDDIPAI SECHRNALKQKEEVCTPYFR
                :*** :*: * *.**:*:*****: ***:**: :*:*: *:*: :*: * * :*
Drome-CYC      FRCKDNSYIQLQSEWRAFKNPWTSEIDYIIAKNSVFL-----
Calfi-CYC      FRAKDGFVKLESKWKQFRNPWTKEIEYLICKNHLLITDEKLKDGGSFVDSGDMNFFSKS
                **.*..*:*:*:*: *:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*
Drome-CYC      -----
Calfi-CYC      SSNSAENSQGNAPMGKDIQRVISSHADAAKIGRRIAEEAAEKFRYDSSASNSPLSLQAAS
Drome-CYC      -----
Calfi-CYC      PRAFIHSRNPSESSEGNARSSADNQLLANAIAEVEQDRSKLCGGVIVSRSSNNCPSTSKG
Drome-CYC      -----
Calfi-CYC      QSVRPEAKNNPNTPGASQPQVNSRGQMSLSEQIMQGAMNARRNSPPSSASSEGND DAA
Drome-CYC      -----
Calfi-CYC      MAVIMSLLEADAGLGGPVDFSGLPWPLP

```

Figure 3.

Putative *Calanus finmarchicus* CYCLE (CYC) protein. Alignment of *Drosophila melanogaster* CYC (Drome-CLK; Accession No. **AAF49107**) with the *C. finmarchicus* CYC (Calfi-CYC) deduced from the Trinity *de novo* transcriptome assembly. In the line immediately below each sequence grouping, “*” indicates amino acids that are identical in the two proteins, while “.” and “:” denote amino acids that are similar in structure between the two sequences. In this figure, helix-loop-helix, PAS and PAC domains identified by SMART analyses are highlighted in yellow, light green and dark green, respectively.

A1. *Drosophila melanogaster* PERIOD vs. *Calanus finmarchicus* PERIOD I

[illegible]

* : * : * : *

Drome-PER
Calfi-PER-I LKELFTELEDGVELEEFVDFCEPTDDDTETSENEHSGIELEEKIFNRNRKKAHLEKMN

Drome-PER
Calfi-PER-I IFMEAEAPFPMDSPRLCVHKHENSQLYGSPYSESTLHLSISGGSEKGSQNSSEDSFKAOK

Drome-PER
Calfi-PER-I KQTKTRFRGDTSSSLGKGSFDKNSGETSHDEDTTENNRVVKFGKKKNVSTITTEFTETKAEV

Drome-PER
Calfi-PER-I EAEKKQDAQI

A2. *Drosophila melanogaster* PERIOD vs. *Calanus finmarchicus* PERIOD II

Drome-PER
Calfi-PER-II MEGGESTESTHTKVSDAYSNSCNSQSRGSSKSLRSGSHSSGSSGYGKPKSTQASS
M-----DEHLGKVNAAAKAHK-DQSEPHNDNVKESEIRSSAFGDSAYGSLGHAYSAY
* : * : * : * : *

Drome-PER
Calfi-PER-II NGPYNLPSGGSFNHLYAKYQNKNTGSSNPSKSSGLSGYMGKQTLVSKNKESKALSNDK
* : : : : * : * : * : *

Drome-PER
Calfi-PER-II EEKPRPSGTGCGVEQQICREL-----DQOQHGEDHSEPAIEQLQQ-EEEEQSGSESEA
RDKSSPQNTKNLKKPEPPLPGFSPSQDTKAARVASLSKALGYIDQIRHKKKKSEKSNEA
. : * : : : : : * : * : * : *

Drome-PER
Calfi-PER-II DRVEGVAKSEAAQSFPIPSPLSVTVIIPP-SMGCGCGVGHAAGLDSGLAKFDKWTWEAGPK
NEL-----IPNQLDVSQTSKLAFPGSESIQHSLG-----SQERVPSN
: : : : * : * : * : * : *

Drome-PER
Calfi-PER-II LESMTGVGAAAAGTQGRSERVKEDSFCCVISMHDGIVLYTTPSIDTVLGYPRDMWLGSRF
L-----GQSEKDEGDFCVAVSLQDGMVHTTNTLSVLGYPKDMWIGRSF
: . : * : * : * : * : * : *

Drome-PER
Calfi-PER-II IDFVHLKDRATFASQITTGPIAESR-----GSVPKDAKSTFCVMLRRYRLGLKSG
IDFVHPMDREVEFISEVTKNVSLPANNMNQSGRPTPGKNITSGKSNDFCRIRRYNLGLKS
***** * . : * : * : * : * : *

Drome-PER
Calfi-PER-II GFGVIGRPVSYEPFRLGLTFREAPEEARPDNMYVNSNGTNMLLVICATPIKSSYKVPDE--
GFAVKERKTKYSFPLKVCVFSEVNTDTSDEDDSDTVPGINFLFIITAIPLQSAISSPNQEG
* : * : * : * : * : *

Drome-PER
Calfi-PER-II --ILSQKSPKFAIRHTATGIISHVDSAASVAGLYPQDLIGRSIMDFYHHEDLSVMKETY
PIVTKCGQEVFISKSTSCIFITFLDENISIFLGYPPKDLQTEIFEHILPNDLIPIKESC
: : : * : * : * : * : * : *

Drome-PER
Calfi-PER-II ETVMKKGTAGASFCSPYRFLIQCNGCYVLETEWTSFVNFWRSKLEFVVGHHRVFGPK
EKIM-----FGHKFSPQYRIKTRNGDFHVHTTVWSCFINFWQQLEFTHGKHTVMRGQ
: * . : * : * : * : * : * : *

Drome-PER
Calfi-PER-II QCNVFEAAPTCCLKIISSEAQSRNTRIKEDIVKRLAETVSRP--SDTVKQEVSRRCQALAS
NPNVFLDPASDCIRNSETHGQNFQIAQAEIKMLQKTVMNRNQLTESIGGAAIKSTKELSS
: * : * : * : * : * : *

Drome-PER
Calfi-PER-II FMETLMDEVSRADLKLPHENELTVSERDSVLMGEISPHHDYDSKSTSTPPSYNQLN
FMGSLGLQVAKAE-NIELSR-----TALAGVQVSPHIDR-----SDSSTPPSYTQLT
* . : * : * : . . . : * : * : *

Drome-PER
Calfi-PER-II YNENLIRFFNSKP-----VTAPAEIDLPKTEPP--EPRGTCVSGASGSPMSVPH
YNENLIRFFNSQPKTLSEKDMLTGTVSPARYEENMKEPQNRKPKETKSAQKKGSGSGHV
***** * : * : * : * : * : *

Drome-PER
Calfi-PER-II EGSGGSGSSGNFTTASNHMSVNTNISIAGTGCTGTGTGTGTGTGTGTGTGTGTGTGT
EDGGRS-----AGSGHSGSGEGHGSGTGGTQTVSGSGDENAKN
* : * : * : * : * : *

Drome-PER
Calfi-PER-II GTGTGTGTGTG-----TGNGTNSGTGTGTASSKSGG-SAAIPVTLTSLLNKH
SSGLTGLSAGHIQPGGEECLMQDGTGTSGLSGSRGEMEDAYKPPALTEELLVTH
. . . : * : * : * : * : * : *

Drome-PER
Calfi-PER-II NDEMEKFLMKHRESRGRTGCK-----SKKSANDTLKMLEYSIPGSGHGK
NRDEKSMILTKFEAK-RIGETRFLLKDGKFKATQDIGSKRVKPDILKLEEK--GDPVR
* . : * : * : * : * : * : *

Drome-PER
Calfi-PER-II RGSHSWEGEANKPKQQLTGLTDAIKGAAGSAGGAVGTGCVSGSGAGVAGGGSGT----
R--KHSSKEPAH--LQITVPNVNKKDQRS-----LAFNDFHWSRRSGMGTPLNI
* : * : * : * : * : *

Drome-PER
Calfi-PER-II -GVAGTTP--EGRATTTSGTGTGGAGGGGGGAGAAAAAGASSVGSSTPGPSSYPTC----
NGISQRFDEHGNMSPAGHIQALPGGSSYQNFQIV--PAYVIEIPHGSGIIPVMLFN
: : . : * : * : * : * : * : *

Drome-PER
Calfi-PER-II TQNLINWPPFVSGITTPPVHSTH-----AMAQSSFSFASAGLPFFFYF-----

Calfi-PER-II	QVPIGLMVDGSTQSFIVPHISDTVAGFNMNPMTEAKASDSESDSLRRTTVHQSKLPNCLAV *.* * .*** :.* * .*. :.* :
Drome-PER	-----IPASLTPTSPTRSPRMHKKHPHGKGTDMPTTSQOAAAAAAQAMP
Calfi-PER-II	NSCMELLQHQEKLQEIATPTSLIKNRLIRLKRSDGSHHTSIKGEPSGALENSASAGASYK :*** . * .*. .*. :. :. :.*. :.
Drome-PER	LQYMAGVMYPHPSLFYTHPAAAAATAMMYQPM-----PPFGMANALQIFE---
Calfi-PER-II	ERFFASNGISEEPVIKETDNECSSQSLYSFTQSSDENDLNDFTPVSSRSRPNGLPEDCK : :.*. . . :. :. :.*. :.*. :.*. :.* :
Drome-PER	-----RPLGSQSAYNKSV-----YTTTFSMTKKVPGAFHVSVTTPAQVQR
Calfi-PER-II	VISPRLVARPLLSEFFWNENVKLDHLIYKQLEEDIEDVLKRDRQFLKMEQPNLVDE *** :.*. :.*.* * .*. :.* :.*. :.* :
Drome-PER	PSSQ-----SASVKTEPGSSAAVSDPCKEVPDSS-----PIPS
Calfi-PER-II	QLTELLADLDFDYQGIDVLFDEDEDSDEETTSADSTGENEDNNKRVKRSKEKVYLEKLN : :. * .*. :.*. :.*. :.*. :.*. :.*. :.* :
Drome-PER	VMGDYNSDPPCCSSSNPANNNKYTDSNGNSDDMDGSSFFSFFYS-SFIKTTDGSSES----
Calfi-PER-II	IFMEENAPFPKSDTNSLNKLTISFGRGLSTVAYGNTSASFKNLDFDTFESSQSSFLAYG : :. :.*. :.*. :.*. :.*. :.*. :.*. :.*. :.*. :.*. :.*. :.*. :.*. :.* :
Drome-PER	-PPDTEKDKPHRKLKMSMTSESKI-----MEHPPEEDQTOHGDC----
Calfi-PER-II	KPSASFKNPDIFDTFENSQSSGVSQNVCTQTIFEKVFENAIEKSSVSRNGHQGDHNSFTS *. :.*. :.*. :.*. :.*. :.*. :.*. :.*. :.*. :.*. :.*. :.*. :.*. :.* :
Drome-PER	-----
Calfi-PER-II	NRPQDKRFRYGGNSSSSSSALEPEKEKNKMDRVSKETKEGKHTSVKTPPEVKSSDNTSSD
Drome-PER	-----
Calfi-PER-II	DTHNLNEQTLKNEVADSRNKKVFPCCERSPHEDHHKIKQDKRFRYGEQASEEKFPVTQKK
Drome-PER	-----
Calfi-PER-II	KEKKRKRKSDNLSKAGYKPTSSPKTRAQDRLKQSTDESSSVYLITSPSAETDPSSD
Drome-PER	-----
Calfi-PER-II	EQTEKKSAAANPDTSDMSISIIISPEKEVKNDNPSEHSDSNISENKETKECGLOQESMTSLD
Drome-PER	-----
Calfi-PER-II	VSGQSGKL

B. *Calanus finmarchicus* PERIOD I vs. PERIOD II

Calfi-PER-I	MGPD-----NQEPEHSGDHCDNDKSLDLSLLFKDSSSEISTQQ--DSAYSQSGQTSTYSLL
Calfi-PER-II	M--DEHLGVNAAAKAHKQSEPHND----NVERKSEISRSASFQDSAYSGLSHAYSNS
	* ** : * : * : * : * : * : * : * : * : * : *
Calfi-PER-I	AANKQPTG-----LHHKTAGTSTGRVSFRSPFKLSKSSSGNSGNSNAALAPPKAV
Calfi-PER-II	GPYNLSPGSGFNHLYAKYQNNKKTGSNP-----SKSSGSLGYMGQTLVQSK--
	.. : * . : * : * : * : * : * : * : * : * : *
Calfi-PER-I	NETTALNLENDNDSRTFRKKKKSGSGSQPPVDESQSGFFNANEKKALQVATLSQALGF
Calfi-PER-II	NEISKALNDKRDKSSPQNTKNLKKESPPLP-----GPFSSQDTKAQVASLSKALGY
	** : * : * : * : * : * : * : * : * : * : *
Calfi-PER-I	VDQFRSHALKIDKANRPVDESEINDLATNLLEASQGA--SVPSTCEFFESLDKQEGISD
Calfi-PER-II	IDQIRH-----KKKSKSEANEALINPQLDVRSTQSLAFPGSESTQHSLSQSERVPS
	: * : * : * : * : * : * : * : * : * : *
Calfi-PER-I	ETFTDSQKKKEGDGFCVAVSLFDGMVHTTSSLTSVLGYPKDMVWGRSFDIFVHPEDRG
Calfi-PER-II	N--LQGSKEDEGDGFCVAVSLQGMVHTTNTLSLGVGYPKDMWGRSFDIFVHPMDREV
	: * : * : * : * : * : * : * : * : * : *
Calfi-PER-I	FITQYTENINLPLRNMVETGKSYDKKELQT--KSGSFFFCRIRIYNGKSGFSIKERKTRYD
Calfi-PER-II	FISEVTKNVSLPANNMQSGRPTPGKNITSKGSNDFFCRIRRYNGLKSGFAVERKRTKYS
	** : * : * : * : * : * : * : * : * : * : *
Calfi-PER-I	PFKLVSVCFEIDKSSSTSKNDVAPQSIYLFITAIPIRSAYSEPFEEGPKVLGTQGTGFI
Calfi-PER-II	SFKLVCVCFSEVNTDTSBDDSDTVGINILFIFTAIPLQSAISSPNQEGPIVTKCGQEVFI
	: * : * : * : * : * : * : * : * : * : *
Calfi-PER-I	SKHTAECKFSMFDEFSISYLGVLQDITGDIDPKYIHPEDMLIKDSFEVIMIEQCKPKPF
Calfi-PER-II	SKHTSSCIPTFLDENSISPLGYFPKDLQGTETFEIHIHPNDLPIIKESCEKIMF--GKHPF
	**** : * : * : * : * : * : * : * : * : * : *
Calfi-PER-I	SKPVRFKINGGYITISSSWSSFINPWSRQLEFVHGKHTVMQGPFPFDIFSEGATLDEQI
Calfi-PER-II	SQPYRKTRNGDGFVHVTTVWSCFINPWSQQLFEIHGKHTVMRGPQNPVFLDPASDCIRN
	* : * : * : * : * : * : * : * : * : * : *
Calfi-PER-I	SDKVYKQGEFLISIEIKCTLKKTLLTNRMNTAISNQSTKNKKELSSFMGSLLEEVAKAENH
Calfi-PER-II	SENTHQGNFAIQAEIKMILQKTVNNRQLTSTIGGAIRKSTKSLSSFMGSLLEQVAKAENI
	* : * : * : * : * : * : * : * : * : * : *

Calfi-PER-I KFGRPNKA--LVIGNISPHYSDSSETPPSYNQLTYNENLTRFFNSQPKTLSEKETLQASG
 Calfi-PER-II ELSRTALAGKVIGNISPHRSDSSETPPSYTQLTYNENLTRFFNSQPKTLSEKDMLTG--
 :.: * :***** :***** :***** :***** : *
 Calfi-PER-I GNTKTASQQFQODSYEGKVQGTSDSKDKHKRAKSTNKSGDGER SANDS--GSGEQ---
 Calfi-PER-II --FVYSPARYEENMKEPQN-RKPKETKSAQKGGSGHVEDGGRSAGSGHSGGQHGSG
 :. : : : : * : : : : : : : : : : * * * * * : *
 Calfi-PER-I -GQTVSGSGEGNGNNSSTLQIGLSAGQNQQGGDGMLSQDGTSGSGSRLSGSGRG--SD
 Calfi-PER-II TGQTVSGSGEDNAKNSSSLQIGLSAGHIQPGGECMLSQDGTSGSGSRLSGSGREGMED
 ***** : : : : : : : : : : * * : : ***** : *
 Calfi-PER-I GYMPPLTEELLAVHNKDMKRLNKFKEAKRTGEMRFLKDSVKQKLAQVAESK---KNI
 Calfi-PER-II AYKPPALTEELLVIHNRDMEKSMITKFKEAKRIGETRFLKDGKFKATQDIGSKRVPKDI
 : * * * * * : : : : * * * * * : * * * * * : * * * * * : * *
 Calfi-PER-I LSKNEELQIRNEKKITKSYKPDQQSHDPNYKSNQVNMDTGMDQPRHFSEAPFGNYIAHN
 Calfi-PER-II IKLKEEKDGPVRRKHSSKPEAHLQITVPD-----NVKKGQDRSLAFNDFHWSS-SRGMG
 :. : * : : * : : : : * : : : : : : : : : : : : : : :
 Calfi-PER-I TTMNTGTMRQVYSGQNHMPFPGG-LVQNTMPQGLHGFTNQFIGVPTVYIPIQGFN----
 Calfi-PER-II TPLNINGISQRDHEGNMSPAGHIIQALPGGSY---QNFIQVPAVYVPIPHGSPV
 * : * : : * : : : : * : : : : * : : : : * : : : : * :
 Calfi-PER-I --PQNAQGIMLQGRGARYVPLGM-----MYSMETVAEEKGLSSKAGQTSQMPELPN
 Calfi-PER-II MLNQVPIGLMVDGTSQSFIPVHISDTVAGFMNPMTEAKASDSEDSLRTTVHQSKLPN
 * : : : * : * : *
 Calfi-PER-I CLAVNSCMVVGQNPPEF-----ARPLSRATSTKGEPSALESNASAS
 Calfi-PER-II CLAVNSCMELLQHQEKLOEIATPTSLIKNRLIRLKSRLGSHHTSIKGEPSALESNASAG
 ***** : : : : : : : : : : * : * : * * * * *
 Calfi-PER-I ASIKNKLGTNASRAQCAGNESEPTAFSTSSSLYSLFKTSD----DF---QSQPSG
 Calfi-PER-II ASVKERFFASNGISEEPVIKTDNE--CSSQSLYSFIQSSDENDLNDFTPVSSRSPNDG
 * : * : : : : : : : : : : : : : : : : : : : * * : : *
 Calfi-PER-I EEEKEKKKKPLKGERPVLSEFFWNQVRKMTSDLIFNYQMETQHLDDVLLKDQALADMQQ
 Calfi-PER-II LPEDCKVISPRLVARPLLSEFFWNENVKLDHLLIYKYLEEEDIEDVLKDRQFLKMEQ
 * . * . * : * * * * * : * : : : * : : : : * : : : *
 Calfi-PER-I PEIVDDQLKELFTELEDGVELEFLVDCEPPTDDDTETSENEHSGIELEEKIFKNRRK
 Calfi-PER-II PNLVDEQLTELLADLDFDYQGIDVLFDEDESDDEETTSADDSTGENEDNNKRVKRSKEK
 * : * : * : * : : : : : : : : : * : * : : : : : : : * : * : *
 Calfi-PER-I AHLEKNIFMEAEAPFMPDSPRLCVHKKENSQLYGSPYYSESTLHSISGGSEKGSQNSE
 Calfi-PER-II VYLEKLNIFMEENAPF-----KSDTNSLNKTDISFKRGLSTVAYGNTSASFKNL
 : * * : * * : * * : * : : : : * : : : : : : : : : * : * : *
 Calfi-PER-I DSKFAGKKQTKTRFGDTSSLGKGSFDKSNGETSHDEDTTENRVIKFGKKKNVSTITEFT
 Calfi-PER-II DLFDTFESSQSFL-----AYGKPSASFKNPDIF---DTFENSQSSGVSQNVCTQTIFE
 * : : : : : * : : * : : * : : * * : * : : * : * : *
 Calfi-PER-I EETKAEVEAEKQDAQ-----
 Calfi-PER-II KVFENAIEKSSVRNQHDNHSFTSNRPQDKFRYGGNSSSSSSALEPEKEKNMDRVSK
 : : : * : : : *
 Calfi-PER-I -----
 Calfi-PER-II ETKEGKHTSVKTPPEVKSSDNTSSDDTHNLNEQTLKNNEVADSRNKKVFPERSPHEDHH

 Calfi-PER-I -----
 Calfi-PER-II KIKQDKRFRYGEQASEEKVPTQKKKEKRKRKRSNLSKAGYPKTSSPKTRAQDPRLKQ

 Calfi-PER-I -----
 Calfi-PER-II STDESESSVEYLITSPSAETDPSSDEQTEKKSANPDTSMSISISIPEKEVKDNPSEHE

 Calfi-PER-I -----I
 Calfi-PER-II SDSNISENKETKECGLQSESTMSLDVSGQSGKL
 :

Figure 4.

Putative *Calanus finmarchicus* PERIOD (PER) proteins. (A1) Alignment of *Drosophila melanogaster* PER (Drome-PER; Accession No. **AAF45804**) with the *C finmarchicus* PER I (Calfi-PER I) deduced from the Trinity *de novo* transcriptome assembly. (A2) Alignment of Drome-PER with *C finmarchicus* PER II (Calfi-PER II). (B) Alignment of Calfi-PER I and II. In the line immediately below each sequence grouping, “*” indicates amino acids that are identical in the two proteins, while “.” and “:” denote amino acids that are similar in structure between the two sequences. In this figure, PAS and PAC domains identified by SMART analyses are highlighted in light green and dark green, respectively. The regions of

Calfi-PER I that corresponds to the PAC regions of Drome-PER and Calfi-PER I is shown in dark green text, though no PAC domain *per se* was identified in this protein using SMART.

		<i>Drosophila melanogaster</i> vs. <i>Calanus finmarchicus</i> TIMELESS
Drome-TIM		MSRVRLHNNHNNQNFQDKVSVMDWLLATPOLYSAFSSLGLEGDTYVVPNALALEE
Calfi-TIM		MD-----NTILNEE-----QILG--NSLGHQVGDYSYVHSETSQVLDN
		*. * * *
Drome-TIM		INRYKLTIEDQTLRTFRRRAIFGQNVRSDLIPLLENAKDD-AVLESVIRILVNLTPVVECL
Calfi-TIM		INRLNEEDKNLRTYRRALAFSQMIQKDLVPILVNSMHNLEVFDAVIRLMVNLTPVVECL
		** : * * * *
Drome-TIM		FSVDVMYRTDVGRHTIFELNKLKLYTSKEAFTEARSTKSVEYMKHILESDPKLSPHKCDQ
Calfi-TIM		CPLIILTKSDVGRQTIYELNSLLKTKEVFRDSRTTRVIMERLSRYATKN-KISSMEVTL
		 * * *
Drome-TIM		INNCLLLLRNLIHIPETHAHCVMPMQSMFPHGISMONTILWNLFQISIDKLLLYLMTCPQ
Calfi-TIM		VNNCLLLLRNLIHIPDQGN-----KNFKQKSSQNMWNLFAQNLDKILLDLIKNNM
		:*****:*****: : : : * * * :***** :*****: *
Drome-TIM		RAFVGVTMVQLIALIYKQDH-----GSGDSSPMLTSDPTSDSS
Calfi-TIM		SSIWCTSVVQLIALVYKQHVITLQKLLQNFLESSESSSESDNESNTSPQYQNGSCSPTT
		:* : :*****:***** : : : * *
Drome-TIM		DNGSNGRGMGGMREGTAATLQEVSRKQ--EYQAMARVPADKPDGSEESDMTGNDSD
Calfi-TIM		DDSSDQASRETSTPELNLAIRERRRSSGSEVLFPNNSPAPSEPMSPGQPMSTSQKS
		*. : * * * *
Drome-TIM		EQPGSPQSQPAGESMDDGDYEDQRHRLNEHGEDE---DEDEVEEEYQLGPASEP
Calfi-TIM		EKSQKRSRQMLKARAARFVNQTNNNETNSSGIDSEIFSTNESEPPQKLYITENEKGEK
		*. * * *
Drome-TIM		LNLTQQPADKVNNTTNTPTSSAPQCLGNEPFPKPPPLPVRASTSAHAQMKFNESSYASH
Calfi-TIM		IVHGRSPNKNKEFVHKMRSAFKN-----PSQLFRSSEPT---EKGSGSSSSG
		: : : : * : : : * * *
Drome-TIM		VSAVKLGQKSPHAGQLQTLTKGKCCPQKRECPSSQSELSDCGYGTQV-----ENQESIS
Calfi-TIM		VIGIPKSSKLPKHKQ-----GSDGSDYGYVTQMEQTMQEHQEMIS
		*. : * * *
Drome-TIM		TSSNDDD-GPQKPKHQKPPCNTKPR-NKPRTIMSPMDKKELRRKLVKRKSSSLINMKG
Calfi-TIM		SSSNDDLLIRNAPPRTNP--VKPRAKDVSKLTPVEKKEHRLKLLKRSKNRMVKA
		:*****: : . : * * *
Drome-TIM		LVQHTPTDDDISNLLKEFTVDVFLKGYSLYVEELHMQLLSN-AKVPIDTSHFFWLVTYFL
Calfi-TIM		MVNHIPITDKDISELLKEFTVDVFLHNGYSLVKELLEKLKNDITMDKSHFLWLTYFL
		:* * * * :*****:***** : * * * * : * *
Drome-TIM		KFAAQLELDMEHIDTILTYDVLSTLYEGVSLCEQLELNARQEGSDLPYLRMRHLVVTA
Calfi-TIM		KFASQLEIGLDQIGTVLSFRTLSYTYEGVELLETLELANRRERRADINPHLRMRHLVVTA
		** : * * * *
Drome-TIM		IREFLQADTYNKVTHLNEDDKAHLRQLQLQISEMSDLRCLFVLLLRFPNPSIHSKQYLQ
Calfi-TIM		LRFEITQVSSYNEIKNLSFSDKQHITKLQSLYTKIRQLQLLLLRFPNPQIQSIQYLS
		:*****:*****: : * * *
Drome-TIM		DLVVTNHILLILDSSAKLGGCQTIR----LSEHITQFATLEVVMHYGILLDEFNNNGE
Calfi-TIM		DLIVCNHMLMDMELVSE---CEEYKGPFSMTNHVQYSNPELMRQYGRLLLEEFPQNT
		** : * * * *
Drome-TIM		FVNDCIFTMHHIGDGLGQIGVLPQPIILKTYRSIWEADYELCDDSDLEIVVIHKFMNT
Calfi-TIM		FVNDCIFTMHHIAGDLDPQVLYIPSILKSFSKIWEQGLQICDDWDVLEIYIIQKFIQT
		:*****:*****: * * *
Drome-TIM		PPKSPLTPTSLTENTKEHNQHTVCSWSQEEMDTLYWYVQSKKNNDIVGKIVKLFESN
Calfi-TIM		MGSSPHSCAANMVECL--DYTEAVDESGFTGAQTSQLFWHFTQVENSKDVPVGAILEVYRQ
		* * * *
Drome-TIM		NGNKLKTRISIIQQLQDDIITLLEYDDLMKFED----AEYQRTLLT-----
Calfi-TIM		TNLSLSRLAVIQALLSHGIITHAQYMFMYKSMVSHKAHEGVSIAEVGSEHEHEHDG
		. * : : : * * * *
Drome-TIM		--TPTSATTESGIEIKECAYGKPSDDVQILLDLIIKENAQHLLWLQRIIECCFVKLTL
Calfi-TIM		HITDSEEPNDSNPEIK----NNNSEIQVLKDCLVKQGRESLIPWLQDVLLDACRVKMYP
		* : * * *
Drome-TIM		RSLGKVPEDGHI-MEPVAYHCICKQKSIPIVQWNEQSTTMLYQPFVLLHLKLGQLPAD
Calfi-TIM		NT--LCPEGSEFPPEPIPFYINKAKQSIPLVFNRIQYQGLQTEAFILMLHLKLGFLPAD
		: * * * *
Drome-TIM		AGSIFARIPDYWTPETMYGLAKKLGPLDKNLKFDASELEDATASSPSRYHHTGPRNSLS
Calfi-TIM		VGVYPRIPHFWSADHIFTVASKIGPIDRNLKFSSTDELEKMS---KHSDETKLSVKN
		* * * *
Drome-TIM		SVSSLDVDLGDTEELALIPEDVAEVAEKAMASTPSPSEIFAVPKTKHCNSIIRTDPDPT
Calfi-TIM		KGEKLPKRASIESAQRKTEIERVENPNNSADTSFLDDMDMNGLESMDLDLLEEPT
		 * * *
Drome-TIM		PPVP----NWLQVMRSKCNHR-TGPSGDPSCVGSSTTVDDDEGFGKSISAATSQAAS
Calfi-TIM		QSKPSTQAGDWLNLVMAKMSRSGSSPSQSSSRKVSSEYSNVMD-----
		. * : : : * * * *
Drome-TIM		TSMSTVNPTTTLNMLNFMGSHNENSSSSGCGGTVSSLSMVALMSTGAAGGGGNTSGL
Calfi-TIM		-----
Drome-TIM		EMDVDASMKSSFERLEVNGSHFSRANNLDQEYSAMVASVYEKEKELNSDNVSLASDLTRM
Calfi-TIM		-MDMD- ** : *
Drome-TIM		YVSEDDRLERTEIRVPHYH
Calfi-TIM		-----

Figure 5.

Putative *Calanus finmarchicus* TIMELESS (TIM) protein. Alignment of *Drosophila melanogaster* TIM (Drome-TIM; Accession No. **AAC46920**) with the *C finmarchicus* TIM (Calfi-TIM) deduced from the Trinity *de novo* transcriptome assembly. In the line immediately below each sequence grouping, “*” indicates amino acids that are identical in the two proteins, while “.” and “:” denote amino acids that are similar in structure between the two sequences. No functional domains were identified by SMART analyses in either protein.

A. *Drosophila melanogaster* vs. *Calanus finmarchicus* CASEIN KINASE II α

Drome-CKII- α	MTLP [*] SAARVYTDVNAHKPDEYWDYENYVVDWGNQDDYQLVRKLG [*] RGKYSEVFEAINITTT
Calfi-CKII- α	MPLASRARVYADVNSHRPREYWDYEAHVVEWGNQDDYQLVRKLG [*] RGKYSEVFESINITNN *.*.* ****:*:*:* ***** :*:*:*****:*****:*****:*****..
Drome-CKII- α	EKCVVKILKPVKKKKIKREIKILENLRGGTNIIITLLAVVKDPVSRTPALIFEHVNTDFK
Calfi-CKII- α	EKCVVKTLKPVKKKKIKREIKILENLRGGTNVITLHG [*] VVKDPVSRTPALIFEHVNTDFK ***** *****:*****:*****:*****:*****:*****:*****:*****
Drome-CKII- α	QLYQTLTDYEIRYYLFELLKALDYCHSMGIMHRDVKPHNVMIDHENRKLRLIDWGLAEFY
Calfi-CKII- α	QLYQTLTDYDIRGYLYELLRALDYCHSMGIMHRDVKPHNVMIDHENRRLRLIDWGLAEFY *****:*:*:*:*:*:*****:*****:*****:*****:*****:*****
Drome-CKII- α	HPGQEYNVRVASRYFKGPELLVDYQMYDYSLDMWSLGCMLASMI [*] FRKEPFFHGHNDYDQL
Calfi-CKII- α	HPGQEYNVRVASRYFKGPELLVDYQMYDYSLDMWSLGCMLASMI [*] FRKEPFFHGHNDYDQL *****:*****:*****:*****:*****:*****:*****:*****
Drome-CKII- α	VRIAKVLGTEELYAYLDKYNIDLDPRFHDILQRHSRKRWERFVHSDNQH [*] LVSP [*] EALD [*] FLD
Calfi-CKII- α	VRIAKVLGTEELYEYLDKYQIELDPRFSDILGRHSRKRWERFVHSENQH [*] LVSP [*] EALD [*] FLD ***** *****:*:*:* ***** *****:*****:*****:*****
Drome-CKII- α	KLLRYDHVDRLTAREAMAHYFLPIVNGQ-----MNPNN
Calfi-CKII- α	KLLRYDHQERLTALEAMEHAYFFPVINDHTRL [*] SAVNSP [*] TGVVSSQATATPASPI [*] LSPGN ***** :***** ***** *.*:*:*:*:*: :.*.*
Drome-CKII- α	Q-----Q
Calfi-CKII- α	SALAGTQPPQ .*

B. *Drosophila melanogaster* vs. *Calanus finmarchicus* CASEIN KINASE II β

Drome-CKII- β	MSSSEEVSWVTWFCGLRGNEFFCEVDEDYIQDKFNL [*] TGLNEQV [*] PNYRQALDMILDLEPED
Calfi-CKII- β	MSSSEEVSWISWFCGLRGNEFFCEVDEDYIQDKFNL [*] TGLNEQV [*] PHYRQALDMILDLEPDD *****:*:*:*:*****:*****:*****:*****:*****:*****:*****
Drome-CKII- β	ELEDNPLQSDMTEQA [*] AEMLYGLIHARYILTNRGIAQ [*] MI [*] EKYQTGDFGHCPRVYCESQ [*] PML
Calfi-CKII- β	ELEDNPNQSDLIEQA [*] AEMLYGLIHARYILTNRGIAQ [*] MI [*] EKYQAGDFGHCPRVYCENQ [*] PML ***** *****:*****:*****:*****:*****:*****:*****:*****
Drome-CKII- β	PLGLSDIPGEAMVKTYCPKCIDVYTPKSSRHHHTDGAYFGTG [*] FPHMLFMVHPEYR [*] PKRPT
Calfi-CKII- β	PIGLSDVPGEAMVKLYCPKCMDVYTPKSSRHHHTDGAYFGTG [*] FPHMLFMVHPEYR [*] PKRPA *:*:*:*:***** *****:*****:*****:*****:*****:*****:*****
Drome-CKII- β	NQFVPRLYGFKIHSLAYQIQ [*] LQAAANFKMPLRAK-N
Calfi-CKII- β	NQFVPRLYGFKIHPLAYQIQ [*] QAAANFKVKPKLKLK *****:*****:***** *****: :*:*

Figure 6.

Putative *Calanus finmarchicus* CASEIN KINASE II (CKII) α and β -subunit proteins. (A) Alignment of *Drosophila melanogaster* CKII α (Drome-CKII α ; Accession No. **AAN11415**) with the *C. finmarchicus* CKII α (Calfi-CKII α) deduced from the Trinity *de novo* transcriptome assembly. (B) Alignment of *D. melanogaster* CKII β (Drome-CKII β ; Accession No. **AAF48093**) with *C. finmarchicus* CKII β (Calfi-CKII β). In the line immediately below each sequence grouping, “*” indicates amino acids that are identical in the two proteins, while “.” and “:” denote amino acids that are similar in structure between the two sequences. In this figure, serine/threonine protein kinase catalytic and casein kinase II regulatory subunit domains identified by SMART analyses are highlighted in purple and black, respectively.

***Drosophila melanogaster* vs. *Calanus finmarchicus* CLOCKWORK ORANGE**

```

Drome-CWO  M-----EPIWNETNGHAAHPVKYES-----
Calfi-CWO  MANTDMSGSSSSNNANTEDMSMDRFQQLSDRSRIDADLQIRIDQNNKMDGLMSKRT
          *          . . . * : : * . . . : : :

Drome-CWO  -----EAAVSSFPYCTESSLNFSTSATAYSEDDAEYATGRNRKTS
Calfi-CWO  DLPDDLACFLQQQKQROARLETTMRSFPTSTE-----AEIEKQEIASKRKRT
          * : : : * * * . * *          : * : : : : * : : :

Drome-CWO  RQDPLSHRIIEKRRRDRMNSCLADLSRLIPPQYQRKGRGRIEKTEIEMAIRHLKHLQSE
Calfi-CWO  HEDPLSHRIIEKRRRDRMNNCLADLSRLIPSSYLKKGRIEKTEIEMAIRHLKHLQSH
          : : * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

Drome-CWO  -CQQKE-----SDYRSGYDCMKEAAKFLYD----VHMQDFCHRLLG
Calfi-CWO  PCNKPEGCEMAQEIETGLSKSNSVESFRVGYHECLTESMHFLVEKEGLYSGNSFCVRLMA
          * : : *          . . . * * * : * : : : * * :          : * * * * :

Drome-CWO  RLQEHIDEMFKTDCYKSTRSCHMPDNVSASSGSPHQAYHPPLCHLRDMLATSASDVEHSQ
Calfi-CWO  HLQKHIDRLGR-----
          : * : * * * : :

Drome-CWO  DHNDVKDLSFRNHLNQLQRSQAAAAAAVAAAAVAVANGSSPASNAGVDSKVPLTNGGGT
Calfi-CWO  -----ASTSEIHSNVGRTWNG--
          * * : : : * : * * . *

Drome-CWO  GGAPPAADNVPSNSTGSGSAAACAGGNSNSGSSNSNAASSTICPPAGGSCPAKVTPLAA
Calfi-CWO  -----QNLPAENGKDRSCERNEMGVKSEYGFSGDESGYSSL-----KTDDDSVNLAK
          : * : : : . * .          * . . . * . * : : : : : : : . * . * *

Drome-CWO  HQQPHQAPVITSTAPHHHHHTDSSHDFESSREPILHTDTSNMHSPPPRDLLLQQHPL
Calfi-CWO  NES-----ERRHEQRNRSYSPEPQSRDDSERGYHYSEQDDKPSDL-----
          : : .          : : * : : : * :          * * :          : . . . * *

Drome-CWO  AHSHTQDSLMSVRMNYSESSHEIEHNNNYKYNHIKERFVHELHDEETSSEHCPVAAH
Calfi-CWO  --SMHGESS-----DNYPKVPEDTSPANQYKFKSNIRHRF--DSMQDADTDY--PPEKR
          * * : . *          * * : . . . . * : * : * : * : * : * : * : * :

Drome-CWO  LQSDHSHLQALSEHSDKGTPEIA---PIMAKKRKLAEAAANGE---IPLEVHTESSNA
Calfi-CWO  QRRNSQPSDKDAERSGNTARSYSPPPTQSHSNYDSFAHLSGHDRMLVPLGRHQLSLSSK
          : : . : : : * : * :          : * : : : : * . * .          : * * * . * .

Drome-CWO  G--ASSANRLDKPSPS--FNFSIDIKIAELHNGNSNSPPLLAKLSAVAAGGQLSTPSS
Calfi-CWO  GDMAPLNHLSEAPDSHYNNNDIK-----QSSSP-----TPYSQGKSFSSPKP
          * * . * : * . : * . : * * * *          . * * *          : . : . * : : * .

Drome-CWO  TTAPLPPTHFTFTVPIFALHGQGNYYVPLNVDYNALVPFLNGMDLLEKSYTSMPPVHPINI
Calfi-CWO  SGVSL-----PIFALNAKGSYYIPLSIDSSVISPFNL--LFADDTYG--PLHPVTI
          : . *          * * * * : : * * * * : * . : * * *          * : . * . : * : *

Drome-CWO  NVNFMPPSSPASLLAAAAAAVAVGKQQ-----QQQAVVAAGAGLPLSTNSAAAQAAAVA
Calfi-CWO  SVNF--TGPTQQPIAPRDQDREREWQRPRNGVIQQPSVIKHWDRAP-----
          . * * : . * : . : * .          : *          * * : :          . *

Drome-CWO  AAVAKAKLEQAMNQSW
Calfi-CWO  -----

```

Figure 7.

Putative *Calanus finmarchicus* CLOCKWORK ORANGE (CWO) protein. Alignment of *Drosophila melanogaster* CWO (Drome-CWO; Accession No. **AAF54527**) with the *C. finmarchicus* CWO (Calfi-CWO) deduced from the Trinity *de novo* transcriptome assembly. In the line immediately below each sequence grouping, “*” indicates amino acids that are identical in the two proteins, while “.” and “:” denote amino acids that are similar in structure between the two sequences. In this figure, helix-loop-helix and orange domains identified by SMART analyses are highlighted in yellow and red, respectively.

A1. *Drosophila melanogaster* DOUBLETIME vs. *Calanus finmarchicus***DOUBLETIME I**

```

Drome-DBT      MELRVGNKYRLGRKIGSGSFGDIYLGTTINTGEEVAIKLECIKTHPQLHIESKFYKTMQ
Calfi-DBT-I     MELRVGNKYRLGRKIGSGSFGDIYLGTTINTGEEVAIKLEGIKTHPQLHIESKFYRIMQ
                *****.*.***** *:*****: **

Drome-DBT      GGIGIPRIIWCSEGDYNVMVMELLGPSLEDLFNFCSSRRFSLKTVLLADQMISRIDYIH
Calfi-DBT-I     GGVGIPAIAKWCSEGDYNVMVMELLGPSLEDLFNFCSSRRFSLKTVLLADQLISREYIH
                **:*** * *****:*****:*****:*****:***:***

Drome-DBT      SRDFIHRDIKPDNFMGLGKKGNLVYIIDFGLAKKFRDARSLKHIPYRENKLTGTARYA
Calfi-DBT-I     SKNFIHRDIKPDNFMGLGKKGNLVYIIDFGLAKKYRDARTHQHIPYRENKLTGTARYA
                *.:*****:*****:*****:*****:*****:*****

Drome-DBT      SINHLGIEQSRDDLESGLGYLMYFNLGALPWQGLKAANKRQKYERISEKKLSTSIIVL
Calfi-DBT-I     SINHLGIEQSRDDMESLGYILMYFVQGTLPWQGLRAATQKQKYERISEKKMSTPIDVL
                *****:*****:***** *:*****:*.:*****:***.*

Drome-DBT      CKGFPSEFVNYLNFRCQMHPDQRPDYCHLRKLFRLNLFHRLGFTYDYVFDWNLKFGGPRN
Calfi-DBT-I     CKGAPCEFATYLNCRSLRFEKPDYSYLRQLFRNLFRHRLGFTYDYVFDWNMLKFGGSRG
                *** *.*.***:***:***:***:***:***** *****:*****.*.

Drome-DBT      PQAIQQAQDGADGQAGHDAAAAAASSHQQQKHVNAALGGGGGSAQQQLQG
Calfi-DBT-I     PET-----
                *:.:

Drome-DBT      GQTLAMLGNGGGNGSQLIGGNLNMDDSMATNSRPPYDTPERRPSIRMRQGGGGGGG
Calfi-DBT-I     -----EADPRY-----
                .: * *

Drome-DBT      GVGVGMPSGGGGGGVGNAK
Calfi-DBT-I     -----ARK
                *

```

A2. *Drosophila melanogaster* DOUBLETIME vs. *Calanus finmarchicus***DOUBLETIME II**

```

Drome-DBT      MSRVNLAVPGPPRDKTPVPKVKGTGKTKDKANPRAKSRSPNPSFCVLSRVKDWE
Calfi-DBT-II    IGAFTPKTALADNVLDLNPSSLPYSEPDQDCSPNNDTDWDWNGKGGGKSGPGSGPAPAN

Drome-DBT      -----MELRVGNKYRLGRKIGSGSFGDIYLGTTINTGEEVAIKL
Calfi-DBT-II    PTENVGLTDKVLSDATTVISDMDIRIAGRYRLGRKIGSGSFGDIYLATDVTNEEVAVKL
                *:***:*****:*****.*.:*.*:***:***

Drome-DBT      ECIRTKHPQLHIESKFYKTMQGGIGIPRIIWCSEGDYNVMVMELLGPSLEDLFNFCSSRR
Calfi-DBT-II    EHSKTKHPQLHVECKFYKVMQGGVGIPLVKYGTGEYNNVMVMELLGPSLEDLFNFCNRK
                *:*****:*.***.***:***:*.:*****:*****:*****.*.:

Drome-DBT      FSLKTVLLADQMISRIDYIHSRDFIHRDIKPDNFMGLGKKGNLVYIIDFGLAKKFRDA
Calfi-DBT-II    LSLKTVLLADQLISRIEFTHGKNFIHRDMKPDNFMGLGKKGNLVYVIDFGLAKKFRDQ
                :*****:*****:***:***:*****:*****:*****:*****

Drome-DBT      RSLKHIPYRENKLTGTARYASINHLGIEQSRDDLESGLGYLMYFNLGALPWQGLKAA
Calfi-DBT-II    RTHQHIPYREHKNLTGTARYTSINHLGIEQSRDDMEALGYILYFLNGTLFWQGLRAK
                *:*****:*****:*****:*****:***:***.*:*****:*.

Drome-DBT      NKRQKYERISEKKLSTSIIVLCKGFPSEFVNYLNFRCQMHPDQRPDYCHLRKLFRLNLFHR
Calfi-DBT-II    TKAQKYEKISEKKLSTPVEELCKGAPAEFATYQNYVRSRLRFEKPDYAYLRQLIRNLFHR
                .* *****:*****.* ***** *.*.*** *:***:***:***:***:*****

Drome-DBT      LGFTYDYVFDWNLKFGGPRNPQAIQQAQDGADGQAGHDAAAAAASSHQQQ
Calfi-DBT-II    QGFSYDYVFDWNTLK-----NGKDPTAQAEIVGADENPPEQKEKE
                **:***** ** *:***.* *.* .....

Drome-DBT      HKVNAALGG-----GGGSAAQQQLQGGQTLAMLGNGGGNGSQLIGGNLNM
Calfi-DBT-II    KKKKAVTGSNQPLDIPTVTGATSAAVTIKATAETQNIIMPVGSSTPQTATTVKLQ-D
                :*:*.*. *.* ***** :. : . . . . * *.*

Drome-DBT      DSMATNSSRP-----PYDTP---
Calfi-DBT-II    KSIISTGNPVPKQIIPPQIINSPPAFLDHNTNYAARVAESATPIPISTHLATDTPVII
                .*:***.* * *****

Drome-DBT      -----ERRPSIRMRQGGGGGGGGVGVG-----MPSGGGGGV
Calfi-DBT-II    LTGCSPLPSQCQPDFNLPAQSLLSLAAPVPTATTSTSQASPASQQLPVMSSG-----
                *:.* * . . . * . . *.*

Drome-DBT      GNAK
Calfi-DBT-II    --LT
                .

```

Comp Biochem Physiol Part D Genomics Proteomics. Author manuscript; available in PMC 2014 September 01.

***Drosophila melanogaster* vs. *Calanus finmarchicus* PROTEIN PHOSPHATASE 1**

```

Drome-PP1 M-----SDIMNIDSIISRLLVGRGAPGKNVQLSESEIRSLCLKSREIFLS
Calfi-PP1 MDSNKVFGQGKVGDTDKLNIDSIIRLLVGRGSRPGKNVQLTENEIRGLCLKSREIFLS
          *          : * :*****:*****:*****:*, ***, *****:***

Drome-PP1 QPILLELEAPLKICGDIHGQYYDLRLFEYGGFPPESNYLFLGDYVDRGKQSLETICLLL
Calfi-PP1 QPILLELEAPLKICGDVHGQYYDLRLFEYGGFPPESNYLFLGDYVDRGKQSLETICLLL
          *****:*****:*****:*****:*****:*****

Drome-PP1 AYKIKYAENFFLLRGNHECASINRIYGFYDECKRRTIKLWKTFTDCFNCLPVAIVDEK
Calfi-PP1 AYKIKYPENFFLLRGNHECASINRIYGFYDECKRRTIKLWKTFTDCFNCLPVAIVDEK
          ***** , ***** , *****

Drome-PP1 IFCHGGGLSPDLSSMEQIRIRIMRPTDVPDQGLLCDLLWSDPKDTMGWGENDRGVSFTFG
Calfi-PP1 IFCHGGGLSPDLQSMQIRIRIMRPTDVPDQGLLCDLLWSDPKDTMGWGENDRGVSFTFG
          ***** , *****

Drome-PP1 AEVVGKFLQKHEFDLICRAHQVVEDGYEFFAKRQLVTLFSAPNYCGEFDNAGAMMSVDDT
Calfi-PP1 AEVVAKFLHKHDFDLICRAHQVVEDGYEFFAKRQLVTLFSAPNYCGEFDNAGAMMSVDET
          **** , ***:***:*****:*****:*****:*****:***

Drome-PP1 LMCSEFQILKPADKRRFVYPNFGSSGRPLTPPRGANNKNNKK-
Calfi-PP1 LMCSEFQILKPADKKKFPYGGL-NAGRPVTPPRGATQQKAKKK
          *****:*** * . : .:***:*****.:***: **

```

Figure 9.

Putative *Calanus finmarchicus* PROTEIN PHOSPHATASE 1 (PP1) protein. Alignment of *Drosophila melanogaster* PP1 (Drome-PP1; Accession No. **CAA39820**) with the *C. finmarchicus* PP1 (Calfi-PP1) deduced from the Trinity *de novo* transcriptome assembly. In the line immediately below each sequence grouping, “*” indicates amino acids that are identical in the two proteins, while “.” and “:” denote amino acids that are similar in structure between the two sequences. In this figure, protein phosphatase 2A catalytic domains identified by SMART analyses are highlighted in dark red.

A1. *Drosophila melanogaster* MICROTUBULE STAR vs. *Calanus finmarchicus*

MICROTUBULE STAR I

Drome-MTS	M--EDKATTKDLLDQWIEQLNECNQITETQVRTLCDKAKEILSKESNVQEVKCPVTVCGDV
Calfi-MTS-I	MNSEKNSLKEGLDQWIEQLMECKQLSENQVKTLECEKAKEILSKESNVQEVKCPVTVCGDV
	* *: * : **: * * * * * *: **: *: *: **: * * * * * : * * * * * * * * * * * * * * * * *
Drome-MTS	HGQFHDLMELFRIGGKSPDTNYLFMGDVYDRGYSVETVTLVALLVKVRFRERITILRGN
Calfi-MTS-I	HGQFHDLMELFRIGGKSPDTNYLFMGDVYDRGYSVETVTLVTLVKVRFRERITILRGNH
	* *
Drome-MTS	ESRQITQVGYFYDECLRKYGANVANWKYFTDLDFYLPALTALVDGQIFCLHGGLSPSIDSLD
Calfi-MTS-I	ESRQITQVGYFYDECLRKYGANVANWKYFTDLDFYLPALTALVDGQIFCLHGGLSPSIDTL
	* *
Drome-MTS	HIRALDRLQEVPEHGPMCDLLWSDPDRGGWGISPRGAGYTFGQDISETFNNNTGLTLVS
Calfi-MTS-I	HIRALDRLQEVPEHGPMCDLLWSDPDRGGWGISPRGAGYTFGQDISETFNNHNGTLTLVS
	* *
Drome-MTS	RAHQLVMEGYNWCHDRNVVTFISAPNYCYRCGNQAALMELDDSLKFSFLQFDPAPRGP
Calfi-MTS-I	RAHQLVMEGYNWSDRNVVTFISAPNYCYRCGNQASIMELDDALKYSYFLQFDPAPRGP
	* *
Drome-MTS	HVTRRTFDYFL
Calfi-MTS-I	HVTRRTFDYFL
	* * * * * * * * * *

A2. *Drosophila melanogaster* MICROTUBULE STAR vs. *Calanus finmarchicus*

MICROTUBULE STAR II

Drome-MTS	M- SDKATTKDLDQWLEQLNECNQLETQV RTLCDKAKEILSKESHNVQEVKCPVTVCGDV
Calfi-MTS-II	MLSMKKEEIASLDKWISQLMERQKLP EMQV KLLCDKAKEILSKESHNVQTVQCPVTVCGDV
	* : * : * : * : * : * : * : * : * : * : * : * : * : * : * : *
Drome-MTS	HGQFHDLMELFRIGGKSPDTNYLFMGD VYDR GYSVETVTL LVAL KVRYRERITILRGNH
Calfi-MTS-II	HGQFHDVLMELFKTGGPLDPTNYLFMGD VYDR GWSVETL LLLA MKVRYRNRITILRGNH
	***** : * : * : * : * : * : * : * : * : * : * : * : * : * : * : *
Drome-MTS	ESRQITQVYGFYDECLRKYG NANV WKYFTDLFDYLPLTALVDGQIFCLHGGLSPSIDSLD
Calfi-MTS-II	ESRQTTQVYGFYDECLQKYGNANAWKYFTDLFDYLPLTALVENKIFCLHGGLSPSLD TL SD
	*** : * : * : * : * : * : * : * : * : * : * : * : * : * : * : *
Drome-MTS	HIRALDRLQ EV PEHGPMCDLLWSDPDDRGWGISPRGAGYTFGQDISETFNN NT GLTLVS
Calfi-MTS-II	HIRALDRIQ EL PEHGPMCDLLWSDV EE CEGWAISPRGAGYLF GKD ISKEFNYSNGLDFVS
	***** : * : * : * : * : * : * : * : * : * : * : * : * : * : * : *
Drome-MTS	RAHQLVMEGYNWCHDRNVVTFISAPNYCYRCGNQAALMELDDSLKFSFLQFDPAPRRGEF
Calfi-MTS-II	RAHQLVMEGYNWAHDKRVVTFISAPNYCYRSGNQAAL LE DDSLNQTFTRFDPATDRGAP
	***** : * : * : * : * : * : * : * : * : * : * : * : * : * : * : *
Drome-MTS	HVTRRTFDYFL
Calfi-MTS-II	SYMNTFDYFL
	: *****

B. *Calanus finmarchicus* MICROTUBULE STAR I vs. II

Calfi-MTS-I	MNSEEKNLSLKELDGWIEQLMECKQLSENQVKTLECKAKEILSKESNVQEVKCPVTVCGDV
Calfi-MTS-II	MLSMKKEEIASLDKWISQLMERQKLPQMQVKLLCDKAKEILSHESNVQTVQCPVTVCGDV
	* *: *: * * * * * : *: * * * * * : * * * * * : * * * * * : * * * * *
Calfi-MTS-I	HGQFHDLMELFRIGGKSPDTNYLFMGDVIDRGYYSVETVTLVTLKVRFRERITILRGNH
Calfi-MTS-II	HGQFHDVLMELFKTGGLPDNTNYLFMGDVIDRGWYSVETLTLALLAMKVRVYRNRITILRGNH
	*****:*****:*****:*****:*****:*****:*****:*****:*****:*****
Calfi-MTS-I	ESRQITQVYGFYDECLRKYGANANWKYFTDLFDYLPALTALVDGQIFCLHGGLSPSIDTLD
Calfi-MTS-II	ESRQTTQVYGFYDECLQKYGNANAWKYFTDLFDYLPALTALVENKIFCLHGGLSPSLDTLD
	*** :*****:*****:*****:*****:*****:*****:*****:*****:*****:*****
Calfi-MTS-I	HIRALDRLQEVPEHEGPMCDLLWSDPDDRGWGISPRGAGYTFGQDISETFNHSNGLTVLS
Calfi-MTS-II	HIRALDRIQLPEHEGPMCDLLWSDVEECGWAISPRGAGYLFKGDISKEFNYSNGLDFVS
	*****:*****:*****:*****:*****:*****:*****:*****:*****:*****
Calfi-MTS-I	RAHQLVMEGYNWSHDRNVVTIFSAPNYCYRCGNQASIMELDDALKYSYLFQDPAPRRGEP
Calfi-MTS-II	RAHQLVMEGYNWAHDRKRVVTIFSAPNYCYRSGNQAALTELDLSLQTFTRFDPATDRGAP
	*****:*****:*****:*****:*****:*****:*****:*****:*****:*****
Calfi-MTS-I	HVTRRTFDYFL
Calfi-MTS-II	SYMNTFDYFL
	:*****

Figure 10.

Putative *Calanus finmarchicus* PROTEIN PHOSPHATASE 2A catalytic subunit
MICROTUBULE STAR (MTS) proteins. **(A1)** Alignment of *Drosophila melanogaster* MTS
(Drome-MTS; Accession No. **AAF52567**) with the *C. finmarchicus* MTS I (Calfi-MTS I)
deduced from the Trinity *de novo* transcriptome assembly. **(A2)** Alignment of Drome-MTS
with *C. finmarchicus* MTS II (Calfi-MTS II). **(B)** Alignment of Calfi-MTS I and II. In the
line immediately below each sequence grouping, “*” indicates amino acids that are identical
in the two proteins, while “.” and “:” denote amino acids that are similar in structure

between the two sequences. In this figure, protein phosphatase 2A catalytic domains identified by SMART analyses are highlighted in dark red.

A1. *Drosophila melanogaster* WIDERBORST vs. *Calanus finmarchicus*

WIDERBORST I

[illegible]

A2. *Drosophila melanogaster* WIDERBORST vs. *Calanus finmarchicus*

WIDERBORST II

[illegible]

B. *Calanus finmarchicus* WIDERBORST I vs. II

Calfi-WBT-I MPGKKAEGKDSSEGGSDSSKAPPVQINKIKYSGPPHIKKERRQSSSRFNITQNRELVK
 Calfi-WBT-II M-----SSGTFVDRIDPFAKRLKK-----KQKKSQGSRRYRTGGDVLT
 ** * : *:: *::*: : **

Calfi-WBT-I LPQLKEATQGEREEELFIQKIRQCGVLFDFVSDPLSDLKWKEVKRAALQEMVEFVTQNRV
 Calfi-WBT-II LPLLKDVPSSEQEELFIRKLQCCVGFDFM-DPVADLKGKEIKRATLNEVVYITAGRGV
 ** *:...*:*****:***** * *:***: ***:**** *:*****:** *:***

Calfi-WBT-I LTEPVYPEVVMFVSFVNLFRITLPPPTNSPSGAEFDPEEDEPTLEAAWPHLQLVYEFFLRCL
 Calfi-WBT-II LTEAVYPEIIRMIACNLFRITLPPSDNP---DFDPEEDDPTLEASWPHLTLVYEFFLRFL
 :**:***: *****:*. * :*****:*****:***** ***** *

Calfi-WBT-I SSDFQPSLAKKYIDQRFVLELLALFDSSEDPREDFLKTTLHRIYKFLGLRAYIRKQINN
 Calfi-WBT-II SPDFQPSIGKKVIDQKFVLQLELEFDSSEDPREDFLKTTLHRIYKFLGLRAYIRKQINN
 *.*****:*. * *:***:***:*** *****:*****:*****:*****:*****

Calfi-WBT-I IF-----YNFIYETERHNGVAELLEILGSIINGFALPLKEEHKIFLLKVLPLH
 Calfi-WBT-II FFLRKVETLGSVKVVFYETEHFNGVGELLEILGSIINGFALPLKAEHKQFLVRVLIPLH
 :* :*:*****:***:*****:*****:***** ***** ** *:***:***:***

Calfi-WBT-I KVRSLSVYHPQLAYCVVQFLEKDPSTLTPVIMSLRLRYWPKVHSPKEVMFLNELEEILDVI
 Calfi-WBT-II KVRCLSLYHAQLAYCVVQFLEKDATLTEQVIKGLLKFWPKTCSQKEVMFLGEIEEILDVI
 ::***:*****:*****:*** ** *:***:***:*** *****:*****

Calfi-WBT-I EPAEFQKVMVPLFNQLTKCVSSPHFQVAERALLYWNNEYIMSLISDNaATILPIMFPALY
 Calfi-WBT-II EPAQFVKIQEPLFKQIAKCVSSPHFQVAERALLYFWNNEYIMSLIEENNQVIMPIMFPALY
 ::***:*****:*****:*****:*****:*****:***:***:*****

Calfi-WBT-I RNSKNHWNTIHLIYNALKLFMEMNQKLFDECTQYKAERMKEKERLQRLDAWNKQVET
 Calfi-WBT-II RISKEHWNQTIVALVYNVLKTFMEMNSRLFDLTASYKADROKKEKRERERDDLWRKMQD
 * **:***:***:***:***:*** *****:***: * :***:***:***:***:***:***

Calfi-WBT-I IARANPDVDQVAKMLDNVSLNSGLPDADMDFDDEATTQEPKTNGANNAGNNINLATVNI
 Calfi-WBT-II LSLKNAAKEGIS--LDMIQMNVA-----
 :: * . : : * : * .

Calfi-WBT-I QEAKKNLDKPLLRRKSDLPTDILTQKSLEYQRADEYLTTPGSDAQNEA
 Calfi-WBT-II -----E

Figure 11.

Putative *Calanus finmarchicus* PROTEIN PHOSPHATASE 2A regulatory subunit WIDERBORST (WBT) proteins. (A1) Alignment of *Drosophila melanogaster* WBT (Drome-WBT; Accession No. **AAF56720**) with the *C. finmarchicus* WBT I (Calfi-WBT I) deduced from the Trinity *de novo* transcriptome assembly. (A2) Alignment of Drome-WBT with *C. finmarchicus* WBT II (Calfi-WBT II). (B) Alignment of Calfi-WBT I and II. In the line immediately below each sequence grouping, “*” indicates amino acids that are identical in the two proteins, while “.” and “:” denote amino acids that are similar in structure between the two sequences. No functional domains were identified by SMART analyses in any of the proteins presented in this figure.

Drosophila melanogaster vs. *Calanus finmarchicus* TWINS

```

Drome-TWS      MGRWRGQSPVLEPPDPQMQTTPPPPTLPPrTFMRQSSITKIGNMLNTAININGAKKPASN
Calfi-TWS      M-----AI-----TGN
                  *                **                :.*

Drome-TWS      GEASWCFSQIKGALDDDDVTDADIISCFEFNHDGELLATGDKGGRVVFQRPASKAANPR
Calfi-TWS      SDIQWCFSQVKGTLDDDDVTEADIISCFEFNQDCELLATGDKGGRVVFQORDAASKSSLPR
                  .: .*****.*:*****:*****.*****.*****.***: **

Drome-TWS      RGEYNVYSTFQSHEPEFDYLKSLIEEKINKIRWLQQKNPVHLLSTNDKTKVLWKVSR
Calfi-TWS      RGEYNVYSTFQSHEPEFDYLKSLIEEKINKIRWLKRKNPAHLLSTNDKTIKLWKVSR
                  *****:*****:*****:*****:*****:*****:*****

Drome-TWS      DKSFGGYNTKEENGLRDPQNVLTALRVPVSVKQIPLVLEASPRRTFANAHTYHINSISVNS
Calfi-TWS      DKRAEYGNLKEDSGLLRDPSSIRALRVPVLRPELMVEASPRRIFANAHTYHINSISVNS
                  **   *** **:.*:***.:.: ***** : : *.***** *****

Drome-TWS      DQETFLSADDLRINLWHLEVNVQSYNIVDIKPTNMEELTEVITAAEFHPTECNVIFYSSS
Calfi-TWS      DQETYLSADDLRINLWHMEITDQSFNIVDIKANMEELTEVITASEFHPKECNLFIFYSSS
                  *****:*****:*****:*****:*****:*****:*****

Drome-TWS      KGTIRLCDMRSAALCDRHSKQFEPEENPTNRSFFSEIISSISDVKLSNSGRYIMISRDYLS
Calfi-TWS      KGTIRLCDMRQAALCDSHKMFEEPDPTNRSFFSEIISSISDVKFSNSGRYIMISRDYLS
                  ***** ***** *** *****:*****:*****:*****

Drome-TWS      IKVWDLHMETKPIETYPVHEYLRAKLCSLYENDCIFDKFECWNGKDSSIMTGSYNNFFR
Calfi-TWS      VKVWDLHMESKPIETFTVHEYLRSLKCSLYENDCIFDKFECWNGTDSAIMTGSYNNFFR
                  :*****:*****:*****:*****:*****:*****:*****

Drome-TWS      VFDRNSKKDVTLEASRDIIKPKTVLKPRKVCTGGKRKKDEISVDCLDFNKKILHTAWHPE
Calfi-TWS      MFDRSSRKEVTLEASREVAKPTMLKPRKVCSSGGKRKKDEISVDCLDFNKKILHTAWHPQ
                  :***.*:.*:*****: *****:*****:*****:*****:*****

Drome-TWS      ENIIA VAATNNLFIFQDKF
Calfi-TWS      ENIIA VAATNNLFLFQENY
                  *****:*****:*****:*****:*****:*****

```

Figure 12.

Putative *Calanus finmarchicus* PROTEIN PHOSPHATASE 2A regulatory subunit TWINS (TWS) protein. Alignment of *Drosophila melanogaster* TWS (Drome-TWS; Accession No. **AAF54498**) with the *C. finmarchicus* TWS (Calfi-TWS) deduced from the Trinity *de novo* transcriptome assembly. In the line immediately below each sequence grouping, “*” indicates amino acids that are identical in the two proteins, while “.” and “:” denote amino acids that are similar in structure between the two sequences. In this figure, WD40 domains identified by SMART analyses are highlighted in teal.

Drosophila melanogaster vs. *Calanus finmarchicus* SHAGGY

[illegible]

Figure 13.

Putative *Calanus finmarchicus* SHAGGY (SGG) protein. Alignment of *Drosophila melanogaster* SGG (Drome-SGG; Accession No. **AAN09084**) with the *C. finmarchicus* SGG (Calfi-SGG) deduced from the Trinity *de novo* transcriptome assembly. In the line immediately below each sequence grouping, “*” indicates amino acids that are identical in the two proteins, while “.” and “:” denote amino acids that are similar in structure between the two sequences. In this figure, serine/threonine protein kinase catalytic domains identified by SMART analyses are highlighted in purple.

Drosophila melanogaster vs. *Calanus finmarchicus* SUPERNUMERARY LIMBS[illegible]

Figure 14.

Putative *Calanus finmarchicus* SUPERNUMERARY LIMBS (SLIMB) protein. Alignment of *Drosophila melanogaster* SLIMB (Drome-SLIMB; Accession No. **AAF55853**) with the *C. finmarchicus* SLIMB (Calfi-SLIMB) deduced from the Trinity *de novo* transcriptome assembly. In the line immediately below each sequence grouping, “*” indicates amino acids that are identical in the two proteins, while “.” and “:” denote amino acids that are similar in structure between the two sequences. In this figure, D domain of beta-TrCP, FBOX, and WD40 domains identified by SMART analyses are highlighted in blue, dark blue and teal, respectively.

***Drosophila melanogaster* vs. *Calanus finmarchicus* VRILLE**

```

Drome-VRI  MSIVCTLEQKVNFFKAAATKNNLLILKNNTDNYINNYKQDNPSNNKFPRIQAQSNNSHL
Calfi-VRI  MKMV-----
          *.:*

Drome-VRI  QHQQQQLQQKLAQLHHYSQQKLSGSDFPYGP RPPTGGKEEKLLLLAPPGLYPEASVSTAM
Calfi-VRI  -----

Drome-VRI  PEVLSGTPTNSHNKANIAMNNVRLSNISPTLSMNGSSNEASNLHPLSMYGGSSISPOSND
Calfi-VRI  -----AEQNLRDQGYAHDIDSRSP
          .**:. . * . .*:.*

Drome-VRI  SGMSDSLKGYVPGSGYGDGMAQSPSQGGNGPQSALTAAQKELFSQRKQREFTPDNKKDE
Calfi-VRI  PTMMLQDTKFFPGH---PASLVQPFSP-----NFSRKDIFNQKQREFIPENKKDD
          . * . .*:.* . :.* * . :*:.*.***** *:****:

Drome-VRI  SYWDRRRRNNEAAKRSREKRRYNDMVLEQRVIELTKENHVLKAQLDAIRDKFNISGENLV
Calfi-VRI  SYWDRRRRNNEAAKRSREKRRMNDMLLETRVIELSKDNHILQAQLNAIFEKYGIKGENMV
          ***** *:.* *****:.*:.*:.*:.* *:.*.***:*

Drome-VRI  SVEKILASLPTSEQVLSNTRAKMSGSGSSGSSSPSGSGSGEGSPQGGHNGYPVGPPLS
Calfi-VRI  SMDVVMSTMPSTNDQVLQFTTK-RLGPMGMRSPSPPLSFSSSLGLPISINNNNNGSNNNN
          *:.*:.*:.*:.*. **: :. * * *. * * * * . :.* . .

Drome-VRI  PLIYGPNGNARPEATVKSVDHIIHAGVAPPPTHLQQLVVPQSQTQHLVQPPQHQPHQQ
Calfi-VRI  NTSVDLNRSCSPERDMSPDARSYR---SPSPTHYH-----
          . * .. ** :. . : : : :*:.*.* :

Drome-VRI  QQISQPPQQQQQQQEPSPSAGSSSPVISDPHNRPPSTTIANLQVQLQALNRNVRPEDLD
Calfi-VRI  ---SSMPQQPL-----ATPSYNDIRSYAESGHMYRRPVADQRYDHEEDRASDKD
          * . *** :.* . * . . * : . * *: :. :.*.*

Drome-VRI  SLRKVVAAGALYNAAAVVGAPPPPSAGLYVPAPSAKDHLEAAAWSHNVEAAVSSSAV
Calfi-VRI  S-----GLALNLSTDRAGSDGTR
          * . * . . : * * . :

Drome-VRI  DAVSSSSVSGSAASVLNLSRRACSPSYEHMLSSTTSSTLSSASSSGAVSGDDEQEHEPAH
Calfi-VRI  DESGRSSASGS-----PPVTSSNIQDGNSSGDESG-----
          * . ***.* . .*****:.*.* **

Drome-VRI  MAPLQLQRSSPQQGSDA---NNCLPLKLRHKSGLGDKDAAATALLSLQH---IKQEPNCS
Calfi-VRI  -----YPRSPNSGSATSGDNSSLPHKLRFSKAVSEKE-AVSSLLSLHQMAMIKREP---
          **:.* : * . * * * . * : : : * :.*.* : **:*

Drome-VRI  RASPAWNDGGDNSSDERDSGISIASAEWTAQLQRKLLAPKEANVVTSAERD-----
Calfi-VRI  -----VEHINHWMETAGNGLTHAGLLASLLPQHSSHFTSIDQDRSEDNGGI
          :. : : . * * * * :.*. :.* :.*

Drome-VRI  -----QMLKSQLERLESEVASIKMILAE-----
Calfi-VRI  EPSRKRLKLPQGIPENIKDEVARTSEVATLKNILVNRMKDEGEDEEREEN
          : :.* : * * * :.* * :.*

```

Figure 15.

Putative *Calanus finmarchicus* VRILLE (VRI) protein. Alignment of *Drosophila melanogaster* VRI (Drome-VRI; Accession No. **AAF52237**) with the *C. finmarchicus* VRI (Calfi-VRI) deduced from the Trinity *de novo* transcriptome assembly. In the line immediately below each sequence grouping, “*” indicates amino acids that are identical in the two proteins, while “.” and “:” denote amino acids that are similar in structure between the two sequences. In this figure, basic region leucine zipper domains identified by SMART analyses are highlighted in pink.

Drosophila melanogaster* CRYPTOCHROME vs. *Calanus finmarchicus
CRYPTOCHROME 1^f

```

Drome-CRY      MATRGANVIWFRHGLRLHDNPALLAALADKD-QGIALIPVFIFDGESAGTKNVGYNRMRF
Calfi-CRY1      MP--GTVIHWFRKGLRLHDNPALIAALKPVNGEVLRLKPVFVLDPWVFNARVGQNRWRF
                *. *: : **:*:*****:*: : : : * **:*:* . . . ** ** **

Drome-CRY      LLDSIQDIDDQLQAATDGRGRLLVFEGEPAYIFRRLHEQVRLHRCIEQDCEPIWNERDE
Calfi-CRY1      LLQSLEDLDSSLKKG---SRLYIVRGKPLETFEDLFHRWDVKRITWEVDTEPYAKERDA
                **:*:*:*:*:*: . ** :.*:* * . *.. ::* * * * :***

Drome-CRY      SIRSICRELNIDFVEKVSHTLWDPQLVIETNGGIPPLTYQMFLHTVQIIGLPPRPADAR
Calfi-CRY1      KVEKVAKDKKIEVVTCTSHTLYNPEQVIAKNKGETPLTYQKMLSLAAIGPPPKPV----
                :...: : :*: * .*****:*: ** * * * .***** :* : ** ***:

Drome-CRY      LEDATFVELDPEFCRSLKLFQ---LPTPEHFNVDNMGFLAKINWRGGETQALLLLD
Calfi-CRY1      ---ETPTELAKEAKASKNDMEKTKYDLPTLKD LGV---NEADLEPCLFPGESEGLNRLA
                * .** * * :*: ** * :.* * * . * : ***:.* *

Drome-CRY      ERLKVEQHAFERGFYLPNQALPNIHDSPKSMSAHLRFGLSVRRFYWSVHDLFKNVQLRA
Calfi-CRY1      DKINMQNASWVRSFEKPNTSPNSLSPSTTVLSPHLKFGCVSARLMYWKLEEVYRGAK---
                :...: : : * * ** : . : *.. :*:*:*:*:* * :*: :...:

Drome-CRY      CVRGVQMTGGAHITGQLIWREYFYTMSVNNPNYDRMEGNDICLSIPWAKPNENLLQSWRL
Calfi-CRY1      ----HSQPPVSLHGQLLWREFFYTVGSNTPNFNMEGNSVCRIQIPWGEDKER-LDAWAE
                : . : **:*:*:*:*: . *.*:*:*:*:*:* * .***** :. * :*:

Drome-CRY      GQTGFPLIDGAMRQLLAEGWLHHTLRNTVATFLTRGGLWQSWEHGLQHFLKYLLDADWSV
Calfi-CRY1      GRTGFPIIDAIMTQLRKEGWIHHLARHSVACFLTRGDLWQSWEAGQHVFEELLLDADWSL
                *:*** ** * ** **:*: * ::* *****.***** * : * : *****:

Drome-CRY      CAGNWMVSSSAFERLLDSSLVTCFVALAKRLDPDGTIYIKQYVPELMNVKPEFVHEPWRM
Calfi-CRY1      NAGNWMWLSASAF--FHQYFRCYSPVAFGKTKDKGAYIRKYLPIFKKFPKAYIYEPWDA
                *****:*:*: : : .*****: * * **:*:*:*: : : * : :*:

Drome-CRY      SAEQQEQYECLIGVHYPERIIDLSMAVKRNLAMKSL-----RNSLITPPPHCRPSNEEE
Calfi-CRY1      PLAVQKQAGCVIGTDYPRPIVKHEVVVKENMAKMKAAAYDAGKANAGKAAPEKRPSEAKE
                . *: * :*:*:* * . . :.*.* ** : * : : * : :*: :*

Drome-CRY      VR-----QFFWLADV V
Calfi-CRY1      PKAKKQKKMDSFFKTEH---
                : .** .

```

Figure 16.

Putative *Calanus finmarchicus* CRYPTOCHROME 1 (CRY1) protein. Alignment of *Drosophila melanogaster* CRY (Drome-CRY; Accession No. **AAC83828**) with the *C. finmarchicus* CRY1 (Calfi-CRY1) deduced from the Trinity *de novo* transcriptome assembly. In the line immediately below each sequence grouping, “*” indicates amino acids that are identical in the two proteins, while “.” and “:” denote amino acids that are similar in structure between the two sequences. No functional domains were identified by SMART analyses in either protein. It should be noted that this protein may be a member of the photolyase family rather than a true CRY1 (see Sections 3.1.3.1 and 3.3 of the text).

***Calanus finmarchicus* CRYPTOCHROME 2 vs. ESTs**

Calfi-CRY2	MTRKTSRRKLTVHWFRKGLRLSDNPALLRGIRRCETFRCIFILDPWFAGSSNVGNTKWR
EL966083	MTRKTSRRKLTVHWFRKGLRLSDNPALLRGIRRCETFRCIFILDPWFAGSSNVGNTKWR
FK041536	MTRKTSRRKLTVHWFRKGLRLSDNPALLRGIRRCETFRCIFILDPWFAGSSNVGNTKWR
FG985765	MTRKTSRRKLTVHWFRKGLRLSDNPALLRGIRRCETFRCIFILDPWFAGSSNVGNTKWR
FK671009	-----
Calfi-CRY2	LLQCLEDLDTSLRKLSRLFVIRGQPADVLPNLFKQWGTTFYTFEEDPEPFGRVRDQNI
EL966083	LLQCLEDLDTSLRKLSRLFVIRGQPADVLPNLFKQWGTTFYTFEEDPEPFGRVRDQNI
FK041536	LLQCLEDLDTSLRKLSRLFVIRGQPADVLPNLFKQWGTTFYTFEEDPEPFGRVRDQNI
FG985765	LLQCLEDLDTSLRKLSRLFVIRGQPADVLPNLFKQWGTTFYTFEEDPEPFGRVRDQNI
FK671009	-----
Calfi-CRY2	AMCKEMNIAVTYEHSHSTLYNLDKIIERNNNKAPLTYKQFQRVIEGMEPPPKVPVTLTQM
EL966083	AMCKEMNIAVTYEHSHSTLYNLDKIIERNNNKAPLTYKQFQRVIEGMEPPPKVPVTLTQM
FK041536	AMCKEMNIAVTYEHSHSTLYNLDKIIERNNNKAPLTYKQFQRVIEGMEPPPKVPVTLTQM
FG985765	AMCKEMNIAVTYEHSHSTLYSLYQIIERNNNKAPLTYKQFQRVIEGMEPPPKACAH-----
FK671009	-----
Calfi-CRY2	LGSAVTPLTDDHDEKYVPTLEELGFDCENLKAPVWTGGETEALSRLERHLERKAWVASE
EL966083	LGSAVTPLTDDHD-----
FK041536	LGSAVTPLT-----
FG985765	-----
FK671009	-----
Calfi-CRY2	GRPKMTPQSLASQTGLSPYLRFGCLSTRLFYHALSDLYRKIKKCDPLSLHGQLLWREF
EL966083	-----
FK041536	-----
FG985765	-----
FK671009	-----
Calfi-CRY2	FYCAATKNPNFDMKSGNPICVQIPWDNNNEALAKWANAHTGFPWIDAIMTQLREEGWIIH
EL966083	-----
FK041536	-----
FG985765	-----
FK671009	-----KWANAHTGFPWIDAIMTQLREEGWIIH
Calfi-CRY2	LARHAVACFLTRGDLWVSWEEGLKVDFELLLDADWSVNAGTWLWLSGSSFFQQFFHCYCP
EL966083	-----
FK041536	-----
FG985765	-----
FK671009	LARHAVACFLTRGDLWVSWEEGLKVDFELLLDADWSVNAGTWLWLSGSSFFQQFFHCYCP
Calfi-CRY2	VRFGRKADANGDFIRKYLPAKNFPTRYIHEPWTAPEAVQSAKCIIGQDYKPKMCHAY
EL966083	-----
FK041536	-----
FG985765	-----
FK671009	VRFGRKADANGDFIRKYLPAKNFPTRYIHEPWTAPEAVQSAKCIIGQDYKPKMCHAY
Calfi-CRY2	VSKLNMERMKQVYAQLAHYRRTGQTGPGIPQDILLSQMKLKPISPFGGLMGPNNPND
EL966083	-----
FK041536	-----
FG985765	-----
FK671009	VSKLNMERMKQVYAQLAHYRRTGQTGPGIPQDILLSQMKLKPISPFGGLMGPNNPND
Calfi-CRY2	SKKGRHLHYDKLASDQTDPDNVQMQYLSNYQIQNLROQGMSPALVQQPSSRSGQPPNMTT
EL966083	-----
FK041536	-----
FG985765	-----
FK671009	S-----
Calfi-CRY2	LQGYTQANRIVPHQRENIILLMDSPGMPQGFIMTSRDQPMFGASTFKSEPSIPSSIINP
Calfi-CRY2	TTTPSIAGSSSGYSSFSNFNSCLSDIKYASSDSLPIVSGDILLPGPISSSTADSIISGAT
Calfi-CRY2	TIVEGCSTISNLSIMSPGSVEHIPSSLSPPVNMGTGSASQENLSSNTPIIGLDNYDAAL
Calfi-CRY2	AEGFFDHSQAQSQN

Figure 17.

Confirmation of protein sequences of the *Calanus finmarchicus* CRYPTOCHROME 2 (CRY2) deduced from the Trinity assembled transcriptome using *C. finmarchicus* expressed sequence tags (ESTs; Accession Nos. **EL966083**, **FK041536**, **FG985765** and **FK671009**). To increase our confidence in the sequences identified as encoding peptide precursors and receptors, each of the deduced proteins was used to query the extant *C. finmarchicus* ESTs curated at NCBI for identical sequences. As can be seen here, BLASTing of the amino acid sequences of Calfi-CRY2 against the extant *C. finmarchicus* ESTs identified four ESTs, each encoding a partial protein homolog. Over the portions of the Calfi-CRY2 sequences

covered by the partial proteins deduced from the ESTs (colored black; the portions of Calfi-CRY2 having no EST overlap are colored gray), all but one amino acid are vetted by at least one overlapping sequence. In this figure, amino acids in the EST-derived sequences that differ from those present in Calfi-CRY2 are colored red; a single uncalled amino acid in the **EL966083**-derived partial protein is noted in green.


```

Drome-CRY      MATR-GANVIWFRHGLRLHDNPALLAALADKDGIALIPVFIFDGESAGTKNVGYNRMRF
GR410803      MSGKEKVTILWFRNGLRLHDNESLLK--ATKDSNSKLLPLFIFDGETPTTKHKYNKMAF
                *: :  .:::***:***** :*  *  *..  *:***:***:  *:  *: *
                .

Drome-CRY      LLDSLQDIDDQLQAATDGRGRLLVFEQEPAYIFRRLHEQVRLHRICIEQDCEPIWNERDE
GR410803      LLECLDDLDLQFW--NQGGKLNLEGGDPVEVLRKISTHETISKLCFDQDCEQIWLERDN
                ***:***: *  :  *: * .***: .:***:  .  :  :***:*** *  ***:
                .

Drome-CRY      SIRSLCRELNIDFVEKVSHTLWDPQLVIETNGGIPPLTYQMFLHTVQIIGLPPRPTADAR
GR410803      AVKNFCGTHKIQIEESIGATLWDPLEIIAANGGSPPLTYQFSHVTKGIGPPRRPIPVDV
                ::::*  :*: *..  ***** :* :*** ***** *  *..  *  *  *  *  .
                .

Drome-CRY      LEDATFVELD--PEFCRSLKLFELPTPEHFNVDNMGFLAKINWRGGETQALLLLDER
GR410803      FSRVQFAKLENFGTLLRSLAVFPTTPTPSMLGIKNSSQ--SEKLYEGGEQRALKYFESR
                . .  *: *  :  *** : *  ***. .: . .  :  :*** : *  : : *
                .

Drome-CRY      LKVEQHA-FERGFYLPNQALPNIHDSPKSMSAHLRFGLSVRRFYWSVHDLFKNVQLRAC
GR410803      MKAERKSLFSRK-----SKPNLEKRSKSLSPYIKFGLSVRKFWALQDGWSQVNKEAP
                :*. *::: *  *  :  ***..  **:*:::*****:***: *  : : *  . *
                .

Drome-CRY      VRGVQMTGGAHITGQLIWREYFYTMSVNNPNYDRMEGNDICLSIPWAKPNENLLQSWRLG
GR410803      PGGSYSIIISHL-----
                *      *

Drome-CRY      QTGFPLIDGAMRQLLAEGWLHHTLRNTVATFLTRGGLWQSWEHGLQHFLKYLLDADWSVC
GR410803      -----

Drome-CRY      AGNWMWVSSSAFERLLDSSLVTCPVAKRLDPDGTIYIKQYVPELMNVKPEFVHEPWMS
GR410803      -----

Drome-CRY      AEQQEQYECLIGVHYPERIIDLSMAVKRNMLAMKSLRNSLITPPPHCRPSNEEEVRFQFFW
GR410803      -----

Drome-CRY      LADV VV
GR410803      -----

```

Figure 18.

Alignment of an expressed sequence tag (EST)-derived putative partial *Calanus finmarchicus* CRYPTOCHROME 1 (Accession No. **GR410803**) and *Drosophila melanogaster* CRYPTOCHROME (Drome-CRY; Accession No. **AAC83828**). In the line immediately below each sequence grouping, “*” indicates amino acids that are identical in the two proteins, while “.” and “:” denote amino acids that are similar in structure between the two sequences. No functional domains were identified by SMART analyses in either protein.

Table 1
Putative *Calanus finmarchicus* circadian protein-encoding transcripts identified by *in silico* transcriptome mining

Query protein	Calanus transcript/protein identifications			
	Transcript		Protein	
	Trinity ID number	E-value	Length [†]	Length [‡]
Core clock proteins				
CLOCK	comp76772_c1_seq1	8.4e-105	1981	Calfi-CLK 540 [*]
CRYPTOCHROME 2	comp181328_c0_seq1	3.5e-227	2690	Calfi-CRY2 794
CYCLE	comp160482_c0_seq1	3.0e-062	2937	Calfi-CYC 685
PERIOD	comp171214_c0_seq1	1.8e-053	4621	Calfi-PER I 1324
	comp205215_c0_seq1	2.4e-050	5198	Calfi-PER II 1581
TIMELESS	comp88114_c0_seq1	1.3e-159	5475	Calfi-TIM 1225
Clock-associated proteins				
CASEIN KINASE II α	comp5221_c0_seq4	3.6e-161	1609	Calfi-CKII α 370
CASEIN KINASE II β	comp606_c0_seq 1	7.4e-107	1507	Calfi-CKII β 216
CLOCKWORK ORNAGE	comp295779_c0_seq1	1.6e-024	2346	Calfi-CWO 587
DOUBLETIME	comp6140_c1_seq1 comp126103_c3_seq2	2.0e-152 9.6e-131	4585 2986	Calfi-DBT I Calfi-DBT II 312 635
PROTEIN PHOSPHATASE 1	comp4558_c0_seq1	3.0e-173	1747	Calfi-PPI 341
PROTEIN PHOSPHATASE 2A-catalytic subunit MICROTUBULE STAR	comp2678_c0_seq3 comp900941_c0_seq1	5.1e-175 6.7e-143	1191 1115	Calfi-MTS I Calfi-MTS II 311 311
PROTEIN PHOSPHATASE 2A-regulatory subunit WIDERBORST	comp25248_c0_seq1 comp18166_c0_seq1	1.0e-142 3.0e-097	3351 4927	Calfi-WBT I Calfi-WBT II 578 480
PROTEIN PHOSPHATASE 2A-regulatory subunit TWINS	comp12910_c0_seq1	4.0e-221	3713	Calfi-TWS 445
SHAGGY	comp8434_c1_seq1	1.8e-179	2513	Calfi-SGG 416
SUPERNUMERARY LIMBS	comp12123_c2_seq1	2.5e-117	2634	Calfi-SLIMB 551
VRILLE	comp71844_c0_seq1	1.6e-035	3139	Calfi-VRI 467
Clock input pathway proteins				
CRYPTOCHROME 1	comp37700_c0_seq1	1.7e-055	1731	Calfi-CRY1 531

[†]Length in nucleotides

*Length in amino acids

* Calfi-CLK is an amino (N)-terminal partial protein. All other proteins listed in this table are putative full-length sequences.

This protein may be a member of the photolyase family rather than a true CRY1.

Table 2

Blastp analyses of putative *Calanus finmarchicus* circadian system proteins vs. all annotated *Drosophila melanogaster* proteins curated in FlyBase and NCBI curated non-redundant protein sequences*

Query	Top FlyBase annotated protein			Top NCBI non-redundant protein		
	FlyBase No.	Species	E-Value	% amino acid identity/ similarity	Accession No.	Species
Core clock proteins						
Calfi-CLK ⁺	FBpp0076500	<i>Drosophila melanogaster</i>	2e-113	23.4/40.7	BAJ16353	<i>Thermobia domestica</i>
Calfi-CRY2	ABA62409	<i>Danaus plexippus</i>	0.0	52.5/74.8	AGA01579	<i>Rhyarobia maderae</i>
Calfi-CYC	FBpp0074693	<i>Drosophila melanogaster</i>	2e-106	31.2/50.1	AB086538	<i>Tribolium castaneum</i>
Calfi-PER I	FBpp0304590	<i>Drosophila melanogaster</i>	4e-080	25.3/54.5	AGA01525	<i>Rhyarobia maderae</i>
Calfi-PER II	FBpp0304590	<i>Drosophila melanogaster</i>	3e-075	22.0/49.3	BAL72155	<i>Apteronomobius asahinai</i>
Calfi-TIM	FBpp0305455	<i>Drosophila melanogaster</i>	1e-127	33.4/65.0	BAJ16356	<i>Gryllus bimaculatus</i>
Clock-associated proteins						
Calfi-CKII α	FBpp0070041	<i>Drosophila melanogaster</i>	2e-161	79.5/87.6	EFN62439	<i>Camponotus floridanus</i>
Calfi-CKII β	FBpp0305406	<i>Drosophila melanogaster</i>	2e-108	82.9/89.3	AAF48093	<i>Drosophila melanogaster</i>
Calfi-CWO	FBpp0081723	<i>Drosophila melanogaster</i>	3e-031	24.8/51.3	EFN67685	<i>Camponotus floridanus</i>
Calfi-DBT I	FBpp0306615	<i>Drosophila melanogaster</i>	2e-155	58.6/67.7	JAA55893	<i>Rhipicephalus pulchellus</i>
Calfi-DBT II	FBpp0306615	<i>Drosophila melanogaster</i>	3e-135	38.1/54.6	JAA55893	<i>Rhipicephalus pulchellus</i>
Calfi-PP I	FBpp0306442	<i>Drosophila melanogaster</i>	2e-177	86.5/94.4	EFN69572	<i>Camponotus floridanus</i>
Calfi-MTS I	FBpp0079148	<i>Drosophila melanogaster</i>	3e-169	92.3/98.1	EEB13988	<i>Pediculus humanus</i>
Calfi-MTS II	FBpp0079148	<i>Drosophila melanogaster</i>	5e-144	76.2/90.7	EEB13988	<i>Pediculus humanus</i>
Calfi-WBT I	FBpp0307570	<i>Drosophila melanogaster</i>	0.0	53.9/71.6	EFN69797	<i>Camponotus floridanus</i>
Calfi-WBT II	FBpp0084575	<i>Drosophila melanogaster</i>	0.0	75.2/84.7	JAA57743	<i>Rhipicephalus pulchellus</i>
Calfi-TWS	FBpp0081671	<i>Drosophila melanogaster</i>	0.0	83.1/96.4	AFK24473	<i>Scylla paramamosini</i>
Calfi-SGG	FBpp0070450	<i>Drosophila melanogaster</i>	0.0	55.1/68.2	EFX72595	<i>Daphnia pulex</i>
Calfi-SLIMB	FBpp0306059	<i>Drosophila melanogaster</i>	0.0	68.8/83.5	EG159076	<i>Acromyrmex echinatior</i>
Calfi-VRI	FBpp0078705	<i>Drosophila melanogaster</i>	1e-037	27.0/50.3	AAT86041	<i>Danaus plexippus</i>
Clock input pathway proteins						
Calfi-CRY I	FBpp0307794	<i>Drosophila melanogaster</i>	1e-145	55.6/79.3	EFX85418	<i>Daphnia pulex</i>

* excluding *C. finmarchicus* proteins, obvious partial proteins, synthetic constructs and provisional protein sequences

[†] Calfi-CLK is an amino-terminal partial sequence

This protein may be a member of the photolyase family rather than a true CRY1.

Table 3

Expressed sequence tag (EST) support for deduced *Calanus finmarchicus* circadian protein sequences

Query	Expressed sequence tags			
	Accession No.	E-value	Protein length*	% identity to query
Core clock proteins				
Calfi-CLK	–	–	–	–
Calfi-CRY2	EL966083	3e-136	193	99.0
	FK041536	3e-133	189	100
	FK671009	4e-131	208	99.5
	FG985765	3e-116	176	94.9
Calfi-CYC	–	–	–	–
Calfi-PER I	FK868108	1e-126	239	94.6
Calfi-PER II	–	–	–	–
Calfi-TIM	ES237746	4e-150	237	95.8
Clock-associated proteins				
Calfi-CKII α	EL696897	9e-144	219	100
Calfi-CKII β	–	–	–	–
Calfi-CWO	FK041142	1e-077	166	82.5 [‡]
Calfi-DBT I	–	–	–	–
Calfi-DBT II	–	–	–	–
Calfi-PP1	FK868144	4e-120	169	100
Calfi-MTS I	–	–	–	–
Calfi-MTS II	–	–	–	–
Calfi-WBT I	EL966219	7e-117	183	98.9
Calfi-WBT II	–	–	–	–
Calfi-TWS	FK041295	2e-121	170	100
Calfi-SGG	FK670737	2e-128	179	100
	FG632527	7e-119	169	100
	EH666828	1e-090	132	100
	EL586005	2e-090	132	100
Calfi-SLIMB	GR410982	6e-086	132	96.2
Calfi-VRI	–	–	–	–

Query	Expressed sequence tags				% identity to query
	Accession No.	E-value	Protein length *	Location on query	
Clock input pathway proteins					
Calfi-CRY1	-	-	-	-	-

Abbreviations: N, amino-terminal partial protein; I, internal fragment; C, carboxy-terminal partial protein.

* In amino acids.

[†]The protein deduced from FK041142 contains a 19 amino acid insertion relative to Calfi-CWO.