Advancing molecular crustacean chronobiology through the characterisation of the circadian clock in two malacostracan species, *Euphausia superba* and *Parhyale hawaiensis*

Thesis submitted for the degree of

Doctor of Philosophy

at the University of Leicester

by

Benjamin James Hunt

Department of Genetics

University of Leicester

2016

Advancing molecular crustacean chronobiology through the characterisation of the circadian clock in two malacostracan species, *Euphausia superba* and *Parhyale hawaiensis*

Benjamin James Hunt

Abstract

The ability to entrain to environmental cycles and therefore anticipate and prepare for the changes they predictably bring is the preserve of the endogenous biological clock, most widely studied at the circadian level. Despite a rich history of research into the behavioural and physiological rhythms shown by many crustacean species, the underlying molecular system driving such traits is not well understood. The aim of this research was to develop our understanding of crustacean clocks through the study of two species, one of major ecological importance and the other a powerful model organism.

The Antarctic krill *Euphausia superba* is a keystone species in the Southern Ocean ecosystem, and evidence suggests that the clock may influence both daily and seasonal rhythms. Using a variety of approaches, including the creation of a *de novo* assembled head transcriptome, a full suite of clock-related genes have now been cloned and characterised. Unlike many species *Euphausia superba* possesses orthologs of every canonical core clock gene, and cell culture assays indicate that the central feedback loop has the capacity for complete transcriptional inhibition via two separate pathways, raising the possibility that the krill clock may be an ancestral type or employ multiple oscillators to control rhythms of differing periods.

In contrast to the relatively intractable krill, the amphipod *Parhyale hawaiensis* has simple maintenance requirements and an extensive genetic toolkit with the potential to enable sophisticated dissection of the molecular clock. With the aim of laying the groundwork for future research the clock genes of this species have also been identified, along with the development of a locomotor activity assay. *Parhyale hawaiensis* shows evidence of bimodal patterns of activity under the control of a molecular clock that combines mammalian-like characteristics with some unique features worthy of further investigation.

Acknowledgements

Contributors

This thesis could not have been produced without the valuable contributions of the following colleagues.

Dr. Ted Gaten:

Collection and storage of *Euphausia superba* samples, as described in section 2.2.1.1.

Entrainment of *Euphausia superba* in LD conditions, subsequent collection and storage of samples in LD and DD for use in RT-PCR, as discussed in section 2.4.3.3.

Dr. Özge Özkaya:

Euphausia superba RNA extraction and storage as described in section 2.2.1.2.

Cloning and sequencing of the initial *Esbmal1* 340 bp fragment through degenerate and semi-degenerate PCR, as detailed in 2.3.1.1.

Tissue-specific PCR of *Esbmal1*, *Escry1* and *Escry2* (Figure 2.7).

Full cloning of *Escry1* and *Escry2* and characterisation of the former as light-sensitive (Figure 2.13).

RT-PCR for *Esbmal1, Esclk, Escry1* and *Escry2* (2.4.3.3; Figure 2.15).

Nathaniel Davies:

Production of the *Euphausia superba* Transcriptome Database, covering the formatting of transcript and annotation data and the production of the website itself. Section 3.2.2.4 details this process in his words.

Table of Contents

CHAPTER 3 THE *EUPHAUSIA SUPERBA* TRANSCRIPTOME DATABASE: *DE NOVO* ASSEMBLY, ANNOTATION AND OUTPUT. 72

List of Tables

List of Figures

Common abbreviations and author's note

AS   Artificial seawater

BCTR   BMAL1 C-Terminal Region

bHLH   Basic helix-loop-helix

CCG   Clock controlled gene

CCID   CLK:CYC inhibition domain

CLD   Cytoplasmic localisation domain

CPR   Caudal photoreceptor

DD   Dark dark

DVM   Diel vertical migration

ERG   Electroretinogram

KSA   Transcriptome created by Konstantinides, Semon and Averof

LD   Light dark

LL   Light light

PAC   C-terminal to PAS domain

PAS   Per-Arnt-Sim domain

PIS   PER interaction site

PRC   Phase response curve

RACE   Rapid amplification of cDNA ends

TMM   Trimmed mean of M values

TPM   Transcripts per million

TTL   Transcription-translation loop

Note: In this thesis there are a number of species that are referenced with such regularity that persistent use of the full or abbreviated binomial nomenclature affects the flow and readability of the text, at least in this author's opinion. The first reference to any species will include the full name, and this will be maintained where deemed necessary for avoidance of ambiguity, particularly in Methods and Results. However, wherever the reader sees only a genus mentioned, it refers to one of the following species.

Euphausia superba; Parhyale hawaiensis; Drosophila melanogaster; Eurydice pulchra; Mus musculus.

Chapter 1 General Introduction

1.1   The biological clock

1.1.1   Ancient rhythms

Almost all organisms experience environmental changes that follow a cyclical pattern, from the 24 hour day-night light and heat cycles, through the phases of the moon and the tides, to seasonal changes in maximum and minimum temperatures and light intensity. Across these rhythmic periods there are predictable times when it is easier to see, or easier to not be seen; times when the danger of dehydration or UV damage is greatly increased, or when food is abundant or scarce. These cycles, if not their current patterns, have been a consistent feature for the entire span of evolutionary history. To respond to them is vital, but given their ubiquity and persistence and the adaptive value to be gained, it is unsurprising that many organisms, from bacteria to bees and moths to mice, show evidence of an endogenous clock that enables their *anticipation*. This is not merely the ability to register and respond to changes, but to internalise the environmental rhythm, so that it persists even in the absence of further external cues.

1.1.2   The benefits of a biological clock

The activity of the biological clock as it relates to 24 hour cycles is known as the circadian rhythm (Latin, *circa* – "around", *diem* – "a day"), the most extensively studied area in the field of chronobiology. To qualify as a circadian rhythm, the activity must meet three criteria: it must persist in constant conditions; it must be possible to entrain or reset it via an external factor such as light; and it must show temperature compensation, continuing unperturbed by changes in environmental temperature (Rosato *et al.*, 1997).

The core advantage of an endogenous circadian clock lies in its ability to entrain to a cycle via an external cue, known as a *zeitgeber*, or "timegiver". Light is the dominant *zeitgeber* for many organisms, but others include temperature (López-Olmeda *et al.*, 2006), barometric pressure (Hayden and Lindberg, 1969) and social cues (Marimuthu *et al.*, 1981). Once entrained, an organism has internalised the passage of time and

uses this information to anticipate environmental changes that predictably follow the cycle. It is thus able to adjust metabolism, physiology and behaviour accordingly, such as foraging at times when food is most abundant, and producing the hormones and enzymes required to make use of that food – the corollary of this is that regulating behaviour and metabolism in this way optimises the use of resources such as time and metabolic substrate, and can also minimise risk from external stresses and dangers (Myers, 2003).

It has been demonstrated that organisms with a circadian clock that closely follows the external cycle have a selective advantage over those that less accurately adhere to it (Ouyang *et al.*, 1998), and circadian dysregulation is linked with a number of human diseases (Hastings *et al.*, 2003). The ability to track time also plays a vital role in other aspects of animal behaviour such as communication (Von Frisch, 1950 cited in Daan (2010) p.6), navigation (Zhu *et al.*, 2008) and reproductive timing, and this latter function has been identified as a mechanism by which speciation may occur through temporal isolation (Tauber *et al.*, 2003).

Organisms exhibit not only circadian rhythms but also ultradian rhythms shorter than 24 hours, circatidal rhythms, and longer circalunar, circasemilunar and seasonal circannual rhythms regulating activities such as reproduction, hibernation and diapause (Ikeno *et al.*, 2010). All are believed likely to be under the influence of the endogenous clock to some extent.

1.2 The molecular clock

1.2.1 The circadian clock in *Drosophila melanogaster*

A great deal of research into circadian biology has been conducted using the fruit fly *Drosophila melanogaster*, a model organism with many advantages including a repertoire of available behavioural assays such as locomotor activity and eclosion timing, and an advanced toolset for genetic manipulation (Jennings, 2011). Consequently, much is known about the endogenous clock of *Drosophila* at the genetic, molecular and neurological level. What follows is an overview of the molecular circadian clock as it is currently understood from studies of this species.

1.2.2 The core

The classic circadian model is based on transcription-translation loops (TTL) (Dunlap and Dunlap, 1999), although there has been a recent shift in focus towards the development of an enzymatic model in which the TTLs regulate and provide robustness to rhythmic enzymatic activity (Lakin-Thomas, 2006; Özkaya and Rosato, 2012). As a major aim of this project is the discovery of the orthologs of the canonical clock genes that underlie the TTLs, this section will cover the mechanics of the classic model.

The basis of the TTL model is a negative feedback loop in which transcription factors drive expression of genes whose proteins feed back to inhibit their own expression through interactions with the original transcription factors (Figure 1.1). At the core of the clock are CLOCK (CLK; Allada *et al.*, 1998; Gekakis *et al.*, 1998) and CYCLE (CYC; Rutila *et al.*, 1998), two basic helix-loop-helix PAS proteins. These form a heterodimer that binds to E-box sequences (CACGTG) and so activates transcription of the clock controlled genes (CCGs) that comprise the output of the circadian system, and also other components of the core clock. Two such components are *period* (*per*; Konopka and Benzer, 1971) and *timeless* (*tim*; Sehgal *et al.*, 1994) – at the protein level, the former is degraded unless stabilised by the latter through dimerisation. The PER-TIM dimer then enters the nucleus and interacts with CLK-CYC to inhibit its own expression. TIM is ultimately degraded by the light-activated flavoprotein CRYPTOCHROME (CRY; Emery *et al.*, 1998), exposing PER to degradation in turn. This allows CLK-CYC transcription activity to restart and the cycle begins again, generating a rhythmic output of CCGs. CLOCK is also involved in a second, less well understood loop, driving the expression of the genes *vrille* and *Pdp1ε* (Cyran *et al.*, 2003). These feed back to respectively negatively or positively regulate transcription of *Clk*, adding stability to the oscillating system.
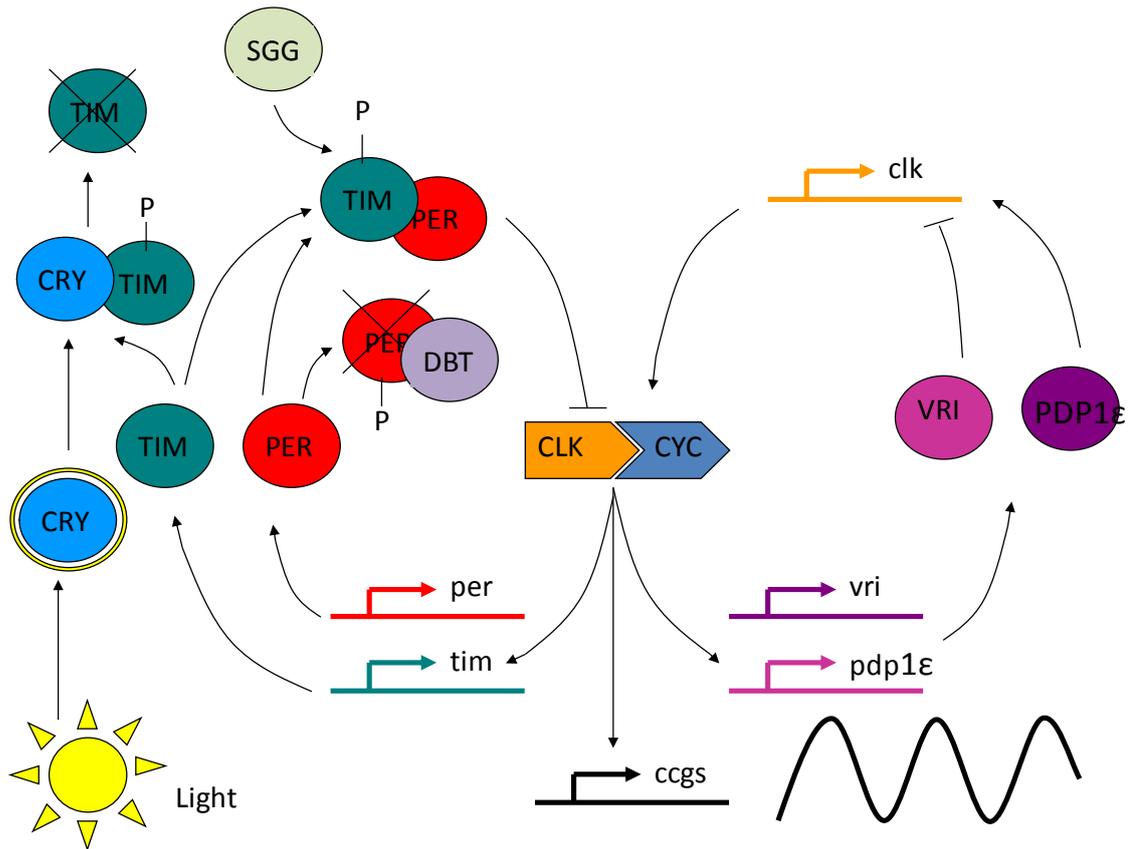
Figure 1.1: CLK and CYC form the central heterodimer, driving expression of *per, tim, vri, Pdp1Ɛ* and other clock controlled genes. PER and TIM, the latter phosphorylated by SGG, feed back to inhibit their own transcription, until degraded by CRY (light-activated) and DBT. VRI and PDP1Ɛ form a second loop respectively repressing or enhancing *Clk* expression. Redrawn from Rosato and Kyriacou (2011).

1.2.3 Supporting roles

The core canonical genes do not run as a standalone mechanism, entraining to external cues and generating rhythms solely through the result of their transcription, translation, inhibition and degradation. Such a cycle would span a few hours rather than the full day required (Gallego and Virshup, 2007), which demands lengthy delays to typical gene expression and protein activity dynamics. A circadian period is instead enabled by the curation of the core feedback loops, which sit at the centre of a complex web of interactions with numerous regulatory components, such as DOUBLETIME and SHAGGY (Figure 1.1). These will be covered in further detail in Chapter Three.

1.2.4 Variations on a theme

The circadian system is remarkably conserved across a wide variety of taxa, with orthologs of each of the core *Drosophila* components and a similar system of feedback loops and post-translational control found in other organisms, albeit with deviations in the details (Reppert and Weaver, 2000). In the mammalian clock, TIM is not involved in forming the repressive dimer, which is instead the role of multiple PER and CRY proteins, the latter of which are not light-sensitive as in *Drosophila*. The monarch butterfly *Danaus plexippus* appears to combine elements of both the mammalian and *Drosophila* system, possessing both a light-sensitive CRY1 and a CRY2 with the ability to abolish the transcriptional activity of the *D. plexippus* CLK:CYC heterodimer (Zhu *et al.*, 2005).

Study of other organisms allows us to build a picture of the ancestral clock. In *Drosophila* the transcriptional activity of the CLK:CYC heterodimer is driven by a glutamine-rich domain in CLK (Allada *et al.*, 1998; Rutila *et al.*, 1998). In mammals and other insects, however, transcriptional activity is due to a transactivation domain in the C-terminus of the CYCLE homolog BMAL1 (Takahata *et al.*, 2000), which appears to be a highly conserved sequence likely to be possessed by the common ancestor of vertebrates and invertebrates, and later lost by *Drosophila* (Chang *et al.*, 2003). Further elucidation of the ancestral clock requires the analysis of a wide range of taxa, beyond mammals and insects. Given that recent phylogenetic analysis places Insecta inside the Pancrustacea clade (Regier *et al.*, 2010), the study of extant crustaceans may prove a productive line of enquiry.

1.3   Crustacean clocks

1.3.1   A history of crustacean chronobiology

Though *Drosophila melanogaster* and *Mus musculus* are dominant in the field due to their multitudinous strengths as model organisms, the subphylum Crustacea has a strong research history in chronobiology. One of the earliest examples of a persistent circadian phenomenon in animals was recorded in the prawn *Hippolyte varians,* changing colour from emerald green to a transparent blue (Gamble and Keeble, 1900). Investigations into the circadian and circatidal rhythms of the fiddler crab *Uca pugnax,*

meanwhile, helped lay the basis for modern chronobiological research with studies on the persistence of chromatophore (Brown *et al.*, 1953) and locomotor (Bennett *et al.*, 1957) rhythms in constant conditions, temperature compensation (Brown and Webb, 1948), persistence of rhythms in isolated leg tissue (Hines, 1954) and the phase-shifting effect of light pulses (Webb, 1950). This latter discovery came well before the concept of the phase response curve (PRC), describing the effect of light pulse timing on the delay or advance of the rhythm phase, was elucidated by De Coursey (1960). Furthermore F.A Brown, from whose lab of much of this work originated, was a strong advocate for the exogenous timing hypothesis to explain circadian phenomena in which the biological clock is held to be guided by an unidentified geophysical force able to surmount the putatively constant lab conditions. This stood in contrast to the endogenous timing hypothesis favoured by Pittendrigh (1960) amongst others, and the conflict between the two ideas drove much productive research (Palmer, 1991).

A wide variety of endogenously controlled, rhythmic phenomena have been identified and studied in crustaceans, at the behavioural, physiological, cellular and hormonal level (Strauss and Dircksen 2010; Table 1.1). Similarly broad is the range of species studied, with representatives from the Classes Malacostraca, Ostracoda, Branchiopoda and Maxillopoda. As the majority of the clade is represented by aquatic animals, many of them marine species, phenomena have also been identified with tidal and lunar/semi-lunar periodicities, as well as sub-24 hour ultradian rhythms (Table 1.2). Studies of these have unveiled a range of associated *zeitgebers* such as hydrostatic pressure, immersion, wave motion and salinity that expand the experimental assays available in the lab.

Table 1.1: Examples of documented circadian phenomena under endogenous control in a range of crustacean species.

Class	Order	Species	Parameter	Reference
Branchiopoda	Cladocera	*Daphnia longispina*	Eye pigment granule migration	Cellier-Michel and Berthon (2003)
		Daphnia pulex	Melatonin synthesis	Schwarzenberger and Wacker (2015)
Malacostraca	Amphipoda	*Talitrus saltator*	Locomotor activity	Bregazzi and Naylor (1972)
			Sun compass orientation	Scapini *et al.*(2005)
		Talorchestia longicornis	Burrow emergence	Forward *et al.* (2007)
	Cumacea	*Dimorphostylis asiatica*	Vertical migration	Akiyama (1997)
	Decapoda	*Chasmagnathus granulata*	Melanophore pigment migration	Granato *et al.* (2004)
		Emerita talpoida	Larval release	Ziegler and Forward (2005)
		Homarus americanus	Molt timing	Waddy and Aiken (2011)
		Nephrops norvegicus	Cardiac activity	Aguzzi *et al.* (2009)
		Procambarus clarkii	Locomotor activity	Page and Larimer (1972)
			Feeding timing	Fernández de Miguel and Arechiga (1994)
			Electroretinogram response amplitude	Verde *et al.* (2007)
		Procambarus spec.	Agonistic behaviour	Farca Luna *et al.* (2009)
		Rhithropanopeus harrisii	Hatching	Forward and Lohmann (1983)
		Scyllarides latus	Locomotor activity	Goldstein *et al.*, (2015)
		Uca pugilator	Distal retinal pigment position	Fingerman (1970)
	Euphausicea	*Euphausia superba*	Vertical migration	Gaten *et al.* (2008)
			Transcript cycling	De Pittà *et al.* (2013)
	Isopoda	*Eurydice pulchra*	Chromatophore dispersal	Zhang *et al.* (2013)
Maxillopoda	Calanoida	*Labidocera aestiva*	Larval release	Marcus (1985)
Ostracoda	Myodocopida	*Asterope mariae*	Vertical migration	Macquart-Moulin (1999)

Table 1.2: Examples of rhythmic phenomena of varying periodicities identified and field and lab studies.

Periodicity	Parameter	Species	Reference
Ultradian	Electrical activity of caudal photoreceptor	*Procambarus clarkii*	Rodríguez-Sosa *et al.* (2008)
Tidal	Swimming activity	*Eurydice pulchra*	Zhang *et al.* (2013)
	Locomotor activity	*Carcinus maenus*	Warman and Naylor (1995)
	Larval release	*Rhithropanopeus harrisii*	Forward and Bourla (2008)
Semi-lunar	Larval release	*Sesarma haematocheir*	Saigusa (1980)
	Courtship	*Uca puligator*	Christy (1978)
	Swimming activity	*Eurydice pulchra*	Alheit and Naylor (1976)
Lunar	Larval release	*Uca* spp.	Skov *et al.* (2005)

The previous paragraphs demonstrate the historical and ongoing contribution of Crustacea to chronobiological research. Where the clade has so far proved lacking, however, is in providing an organism with a toolkit of molecular techniques comparable to those employed to dissect and locate the source of the circadian clock at the neurological level in *Drosophila melanogaster*. This is an issue that is addressed further in section 1.6.

1.3.2 Molecular data: a long deficit, then a sudden deluge

Until recently knowledge of crustacean clock genes was sorely lacking, with a single gene characterised in *Macrobrachium rosenbergii* (Yang *et al.*, 2006) up until the beginning of 2010. A full suite of putative circadian genes was identified through searching the genome of *Daphnia pulex* (Tilden *et al.*, 2011) and in the last few years the falling cost and increasing throughput of next generation sequencing technology such as RNA-seq has improved matters greatly, with suites of putative circadian genes identified from the transcriptomes of the copepods *Tigriopus californicus* and *Calanus finmarchicus* for example (Nesbit and Christie, 2014; Christie *et al.* 2013a). Much work remains to be done in characterising and confirming the function and activity of such genes, however; at this current time we have an abundance of gene coding sequences and little knowledge of what they do, and where, and how. A more thorough study has been made of the circadian and circa-tidal clock of the isopod *Eurydice pulchra* (Zhang *et al.*, 2013), in which the core clock genes have first been cloned, then assessed for transcript cycling and, through the use of *Drosophila* embryonic S2 cell assays, characterised in their interactions with each other to build a model of that animal's clockwork. It was also shown that *Eurydice*'s circa-tidal clock is at least partially independent of its circadian clock, with tidal swimming behaviour unaffected by constant light that abolished the observed circadian chromatophore rhythm, or by RNAi knockdown of *Eurydice period* expression, which again dampened chromatophore rhythms.

The Antarctic krill *Euphausia superba*, a keystone species of the Southern Ocean ecosystem, shows a diel vertical migration pattern (Godlewska and Klusek, 1987) that may be under the influence of an endogenous clock, but as yet only a single circadian gene has been published (Mazzotta *et al.,* 2010). One of the aims of this project is to

expand our knowledge of the core clock genes of *Euphausia superba,* not only obtaining genetic sequence data but also examining gene expression and protein interactions; in short, to build a model of the krill clock.

1.4 *Euphausia superba*

1.4.1 Distribution and biomass

Euphausia superba (Dana, 1852; hereafter *Euphausia* unless otherwise specified) is a small pelagic crustacean of the order Euphausiacea that inhabits a circumpolar belt in the Southern Ocean between the Antarctic continent and the Polar Front (Figure 1.2). It is hugely abundant within these waters with a biomass estimated between 125 and 725 million tonnes, but the distribution is not uniform and there are regions of high population density in areas such as the Scotia Sea and the Western Antarctic Peninsula (Murphy *et al.*, 2007). *Euphausia* is a highly social organism that exhibits strong swimming ability and swarming behaviour, gathering in dense schools ranging from a few m^2 to superswarms covering kilometres of ocean and depths of over 100 m (Tarling *et al.*, 2009). Adult krill tend to occupy the top 200 m of the water column, but have been recorded present in significant numbers feeding at depths of over 3 km (Clarke and Tyler, 2008).
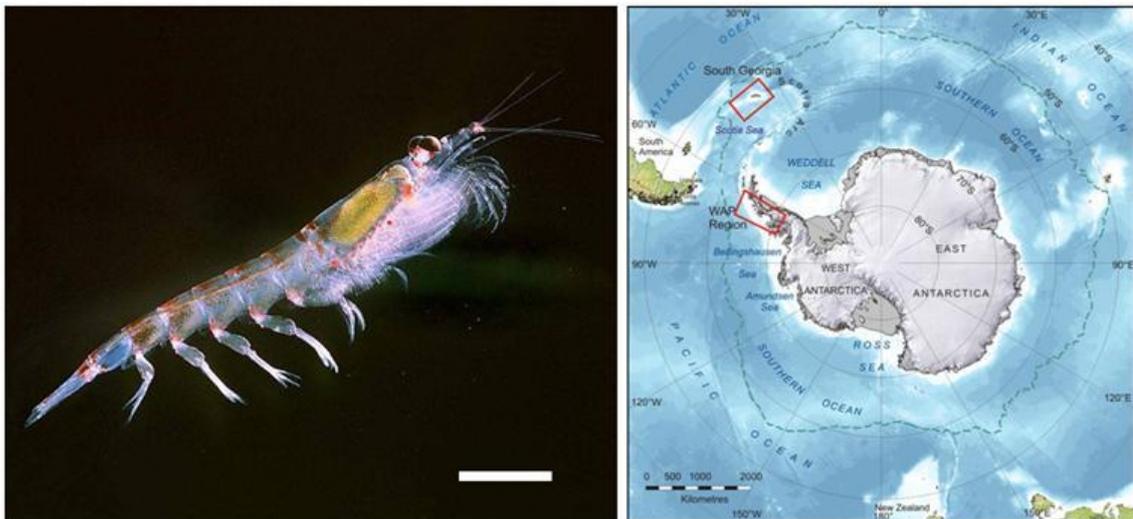
Figure 1.2: Left - adult *Euphausia superba*, scale bar = 1 cm (edited to add scale bar, Wikipedia, 2009). Right - the Southern Ocean. Dotted line delineates the Polar Front, red boxes indicate areas of high *Euphausia* population density (Murphy *et al.*, 2013).

1.4.2 Life history

Reproduction occurs during the summer months, with females able to spawn multiple times in one season and over multiple seasons (Ross and Quetin, 1986) and producing an average of 2,500 eggs per spawning (Ross and Quetin, 1983), which are fertilised by the attached male spermatophore upon release. The eggs sink and are estimated to hatch 6 days later at a depth of around 850 m (Quetin and Ross, 1984). The larvae have a complex developmental sequence comprised of three naupliar stages, three calyptopis stages and six furcilia stages before emergence as a juvenile (Ross and Quetin, 1986), during which time they will overwinter under the Antarctic pack ice, which provides food and refuge from predators (Frazer *et al.*, 2002). *Euphausia* continue to moult regularly throughout adulthood, and retain the ability to grow or shrink after each moult, the latter of which may occur in response to poor food availability and which may play a role in overwinter survival (Alonzo and Mangel, 2001). As a result of this variable adult size and the seasonal loss and recovery of secondary sexual characteristics, an accurate picture of the overall population structure is difficult to assemble (Nicol, 1990).

Euphausia are relatively long-lived, field observations suggesting a lifespan of up to 7 years (Ettershank, 1983) and are able to survive long periods of starvation – over 200 days in laboratory settings (Ikeda and Dixon, 1982). It is an omnivorous, raptorial organism (Price *et al.*, 1988) that feeds in the main part on phytoplankton, but also makes use of sea ice algae, zooplankton and benthic detritus (Murphy *et al.*, 2007).

1.4.3 Ecological and economic importance

These life history traits – biomass, lifespan, omnivory – combine to make *Euphausia* a keystone species in the Southern Ocean ecosystem, the main link between primary production and higher trophic levels, comprising up to 70% of the food intake of predators such seals, seabirds, whales, squid, fish and penguins, with an annual demand of up to 25 million tonnes (Murphy *et al.*, 2007). The ability to undergo long periods of starvation allows *Euphausia* to function as a link through areas of low productivity to enable higher level consumption far from sites of high productivity. There is evidence that some *Euphausia* populations may have declined dramatically in recent decades, due in part to changing temperatures in the Antarctic and the

resulting decline in sea ice extent (Atkinson *et al.*, 2004), and a concurrent decline in certain krill-dependent predators has been observed (Reid and Croxall, 2001).

Furthermore, *Euphausia* has been suggested to play the role of 'ecosystem engineer' in the Southern Ocean (Murphy *et al.*, 2013), defined as an organism that can change the abiotic environment through their feeding, grazing and excretion activities, increasing or decreasing the availability of resources to other organisms to the extent that they play a significant role in defining the structure of the ecosystem (Jones *et al.*, 1994; Jordan, 2009). By dint of its biomass, omnivorous nature and production of large, dense, rapidly sinking fecal pellets, *Euphausia* plays an important role in the phenomenon known as the biological pump in which carbon is removed from the ocean surface and sequestered at depth for hundreds of years (Perissinotto *et al.*, 2000; Tarling and Johnson, 2006). It has also been linked with iron recycling from the seabed to the surface through its benthic feeding habits (Schmidt *et al.*, 2011) and production and uptake of ammonium, the preferred nitrogen source of phytoplankton (Whitehouse *et al.*, 2011).

Euphausia superba is of growing economic importance with a rising but still under-exploited fishery (Nicol *et al.*, 2012). The main commercial use of *Euphausia* derived products is in aquaculture, where it is used both as a bulk ingredient in fish meal and as a nutritional additive, but there is also a growing demand for products for human use in the realms of both pharmaceutical and complementary medicine due to *Euphausia's* high omega-3 polyunsaturated fatty acid content, dietary intake of which has been linked with many beneficial health effects (Hooper *et al.*, 2004).

1.5 The rhythms of *Euphausia superba*

1.5.1 Timekeeping in the Southern Ocean

The dominant *zeitgeber* is often light, and changes in photoperiod bring adjustments to the circadian rhythm as the organism entrains to the new cycle. In constant light conditions, *Drosophila* becomes arrhythmic – the endless light stimulus essentially crashes the circadian clock. In constant darkness, endogenous rhythms surface that can show different periodicity to the external cycle, a product of the organism's internal timekeeping system.

At the North and South pole both constant light (summer) and constant darkness (winter) are experienced as part of the natural seasonal cycle, and light quality can be hugely variable over the course of a single day. For *Euphausia* this situation is further complicated by their aquatic habitat, given that much of the light spectrum does not penetrate far below the surface of the ocean and they overwinter under extensive sea ice. This raises the question as to whether light could be the dominant *zeitgeber*, and if this is not the case it will be necessary to determine what is; hydrostatic pressure, feeding and social behaviour have been suggested as candidates (Gaten *et al.*, 2008).

A key reason, then, for the interest of chronobiologists in polar organisms is to ascertain if, and how, they maintain rhythmic behaviour and physiology in the face of conditions that offer such challenges to the biological clock. For example van Oort *et al.* (2005) found that Arctic reindeer *Rangifer tarandus platyrhynchus* and *Rangifer tarandus tarandus* showed rhythmic behaviour when there was a clear light/dark cycle that was absent during the constant conditions of summer and winter. The freshwater Arctic branchiopod *Lepidurus arcticus* appears to show no circadian rhythmicity in locomotor activity, albeit on the basis of short recordings of a small number of animals (Pasquali, 2015). Berge *et al.* (2009), on the other hand, recorded acoustic data showing that the diel vertical migration (DVM) of Arctic zooplankton continued throughout the winter months, although in accounting for the data an endogenous pacemaker was rejected on the basis of a brief pause in the pattern, the authors instead invoking lunar and low-level solar illumination to explain its persistence – this conclusion is supported by the recent findings of Last *et al.* (2016), which is discussed further in Chapter Six. Looking to *Euphausia,* it undoubtedly recognises and responds to the daily and seasonal rhythms of the Southern Ocean with changes in behaviour, physiology and metabolism; as the Antarctic climate continues to change, it will be important to determine if these phenomena are under the control of an endogenous clock so as to be able to predict how *Euphausia* will adjust, and how this will affect the ecosystem as a whole.

1.5.2   Circadian behaviour

Like many marine organisms, *Euphausia* changes its position in the water column over the course of the day, exhibiting a DVM pattern in which daylight hours are spent at

depths of around 100 m, rising to the surface at night and returning to deeper waters at dawn (Figure 1.3). It is thought that this enables the exploitation of the more abundant food resources present at the surface while minimising the risk of predation by visual predators (Lampert, 2007; Gliwicz, 1986), and may also play a role in minimising the damage from UV radiation (Auerswald *et al.*, 2008).

Although a widely studied phenomenon in a number of marine crustaceans, documented instances of DVM under the influence of an endogenous pacemaker – an entrained pattern with persistence in constant conditions - are rather rarer, no doubt due to the logistical challenges of conducting controlled, lab-based experiments; many organisms exhibiting the behaviour are swarming, pelagic and fragile. Harris (1963) demonstrated the endogenous nature of vertical migration in the copepod *Calanus finmarchius* and the cladoceran *Daphnia magna*, the animals showing a pattern of upwards migration during the dark hours and sinking during the day that persisted in constant darkness. A study of the DVM of various zooplankton by Enright and Hamner (1967) identified persistence in constant conditions in *Nototropis* amphipods and Peltidiad copepods.
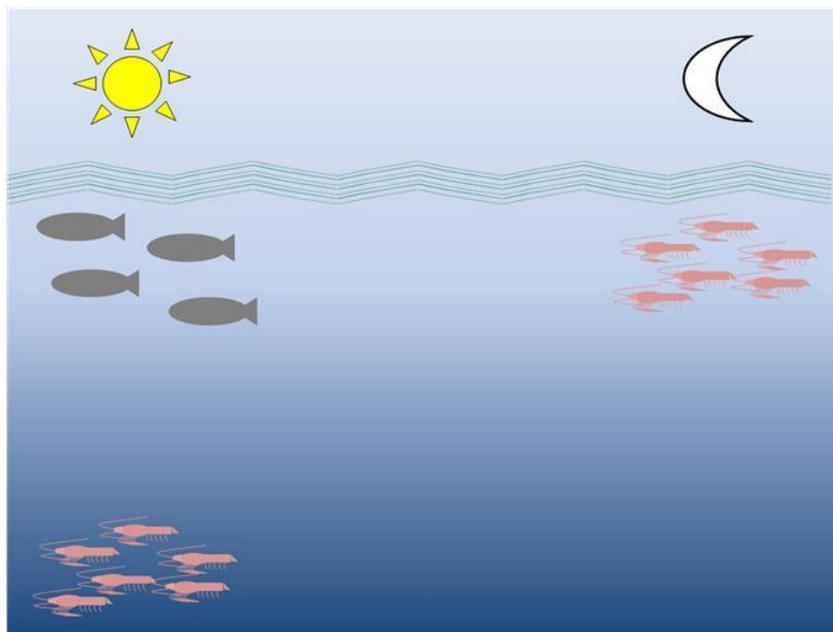
Figure 1.3: Diel vertical migration. *Euphausia* remain ~100 m below the surface during the day to avoid visual predators. At night they rise to the surface to exploit the more abundant food resources there.

Environmental factors such as the presence of food and predators, light intensity, social interactions and photoperiod no doubt play a role in controlling the DVM of *Euphausia*; it is more variable than that observed in some other planktonic species, with additional upward migrations during daylight hours that appear to be influenced by hunger and equivalent sinking during the night when satiated (Tarling and Johnson, 2006), and seasonal changes in its robustness (Taki *et al.*, 2005). This suggests a complex system with multiple influences that is difficult to characterise at the population level. But given the benefit derived from the ability to anticipate environmental changes it is probable that it is also regulated by an endogenous clock, allowing upward migration to begin before sundown to maximise feeding time, and downward migration to begin before sunrise to avoid the proverbial and literal 'early bird'.

This view is supported by Gaten *et al.* (2008), a study that investigated the role of an endogenous circadian clock in *Euphausia* DVM at the individual level. Single specimens were isolated in 'activity monitors', cylinders containing 5 L seawater with infrared barriers near the top and bottom that registered whenever the barrier was broken by the movement of the krill – a clear DVM pattern in these cylinders would show as a high number of responses at the top barrier during 'night' and vice versa during the 'day'. The krill were kept first in LD (light-dark) conditions that mimicked the natural light cycle for 5 days to observe the 'normal' activity pattern, if any, and then switched to DD (continuous darkness) for another 5 days to ascertain if this pattern was maintained in constant conditions, which would suggest the presence of an endogenous clock. The results showed that, as with previous field observations, *Euphausia* has a complex activity pattern with multiple rhythmic components, but mean activity tended to be higher at the top sensor during the 'dark' periods of the LD and DD experiments and higher at the bottom sensor during the 'light' periods, suggesting that there is a rhythmic activity pattern in *Euphausia* that is maintained in constant darkness. The complexity of the pattern lead the authors to suggest *Euphausia* may possess a multi-oscillator system with the ability to annihilate overt rhythmic behaviour during periods of constant darkness or light, when this could be

non-beneficial or even detrimental; not the first time that such a multi-oscillator model has been suggested for crustacean clocks (see section 1.6.3). *Euphausia* did not appear to entrain to light during the LD component of the experiment, suggesting light may not be a dominant *zeitgeber* in krill.

Bimodal circadian oscillations in oxygen consumption, metabolic enzyme activity have also been documented in *Euphausia* kept in aquaria, as well as rhythmic gene expression of a key component of the circadian clock (Teschke *et al.*, 2011), and hundreds of other genes have been to shown to be rhythmically expressed through the production of a diurnal transcriptome (De Pittà *et al.*, 2013).

1.5.3   Seasonal variations

Rhythms over periods greater than 24 hours are also seen in *Euphausia*. Thomas and Ikeda (1987) showed that mature female *Euphausia* undergo a moult-shrink regression process to a juvenile state over winter and then return to maturity in spring, a trait that has been found to persist in constant conditions and thus may be under control of a circannual clock (Kawaguchi *et al.*, 2007). Metabolic rate also follows a seasonal cycle, falling in the autumn and through winter and increasing once more in spring (Meyer *et al.*, 2010).

Seear *et al.*, (2009) have found a possible link between such seasonal cycles and the clock. *Euphausia* were captured and kept in either LD (12:12) or DD winter-simulation conditions for 7 days, and then total RNA extracted and analysed for differences in gene expression between the two regimes. LD krill expressed genes involved in moulting and metabolism at a significantly higher level than those kept in DD, in line with the findings discussed above. Of particular interest is that these differences were triggered by exposure to different photoperiods: Hirano *et al.* (2003) found that light could trigger faster maturation of juvenile krill and in many organisms it has been shown that changes in photoperiod govern seasonal developments via the components of the circadian clock (Ikeno *et al.*, 2010; Yanovsky and Kay, 2003). Further research will show if *Euphausia*'s response to a seasonal light regime is intertwined with their daily rhythms.

1.6   On the productivity of model and non-model organisms

1.6.1   *Euphausia superba*: high importance, high maintenance

One of the aims of this project is to add to what is currently known of the crustacean clock; how it works, what aspects are shared with other species, how it differs, and what this might tell us about the evolution of the animal clock. And while it is inarguably a creature of vital importance to the Southern Ocean ecosystem and a fascinating animal in its own right, if a vote were conducted to crown the best model organism for lab research *Euphausia superba* would be unlikely to mount a strong challenge. It is a fragile, highly social animal that breeds only during the summer (Ross and Quetin, 1986), lives for upwards of five years (Ettershank, 1983) and has complex aquaria requirements, the fulfilment of which still does not prevent high mortality and low spawning rates (Hirano *et al.*, 2003). Isolation breeds stress, making behavioural studies of individuals unrepresentative of natural behaviour, and yet they also show great resistance to schooling in lab aquaria (Ritz *et al.*, 2003) at which point the natural behaviour of the researcher is to throw their hands up in frustration. Collection of samples involves spending months on a trawler in the Antarctic and is particularly expensive and challenging should one wish to collect during winter. It has an enormous genome of approximately 47 Gbp (Jeffery, 2012), unsequenced and likely to remain so for the foreseeable future. In short, what *Euphausia* can teach us about the circadian clock, even just that of crustaceans, is limited by its intractable nature.

1.6.2   The very model of a modern research animal

Contrast that with *Drosophila melanogaster.* A small, hardy and highly fecund fly with a short generation time, it breeds quickly and year-round and is happy to subsist on an agar-based mix of yeast, starch and sugar that doubles as an egg-laying substrate. It is in possession of one of the most comprehensive molecular toolkits in research allowing the generation of targeted mutations, inducible knock down or over-expression of genes in the whole organism or individual tissues, easily visualised *in vivo* reporters and much more (Jennings, 2011). In stark contrast to *Euphausia*, the euchromatic portion of its sequenced genome was reported at 120 Mb (Adams *et al.*, 2000).

These powerful molecular techniques have allowed the genetic and neural basis of the circadian clock in *Drosophila* to be comprehensively described. With increasing precision via antibody staining, analysis of mutant strains and cell-specific expression of reporter genes the source of the clock – the pacemaker that sets the rhythm for the rest of the body – was determined to lie in approximately 150 neurons mainly projecting to the dorsal protocerebrum (see the review by Helfrich-Förster (2005) and Figure 1.4). These are classified into either lateral or dorsal neurons based on their position in the brain, each type further subdivided into 3 clusters; the lateral neurons comprise six lateral dorsal neurons (LN$_d$), five large lateral ventral neurons (l-LN$_v$) and five small lateral ventral neurons (s-LN$_v$). The dorsal neuron clusters are named simply DN$_1$ (roughly 15 cells), DN$_2$ (two cells) and DN$_3$ (roughly 40 cells).
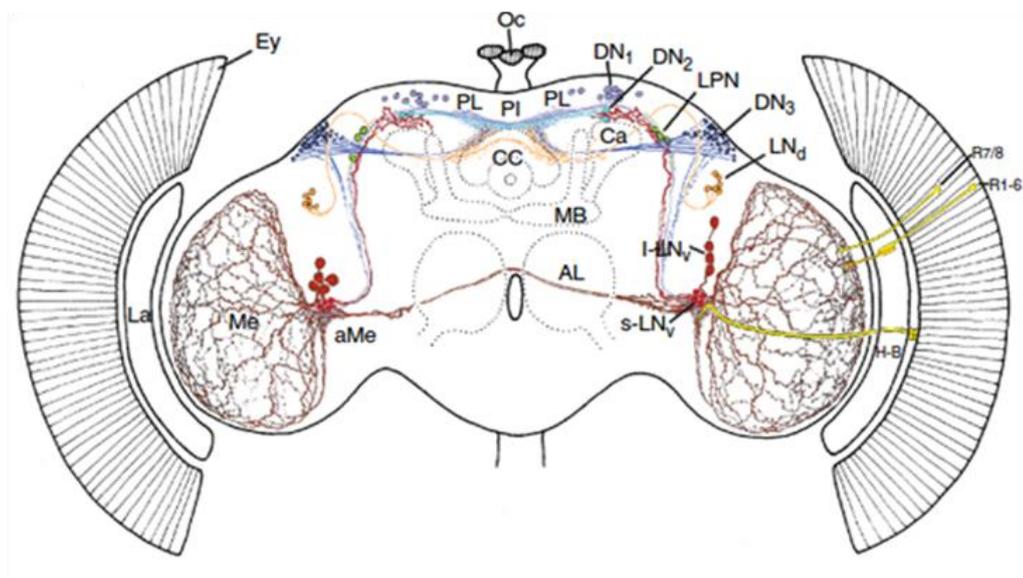
Figure 1.4: Clock neurons in the *Drosophila* brain. Ey – compound eye. La – lamina. Me – medulla. aMe – accessory medulla. PL/PI – pars intercerebralis/lateralis. CC – central complex. MB – mushroom bodies. AL – antennal lobe. Ca – calyces. Oc – ocelli. R1 – R8 – photoreceptor cells. H-B – eyelet cells. All other labels – see text. From Helfrich-Förster (2005)

Of these clusters, the most important in generating a sustained rhythm appear to be the LN$_d$ and s-LN$_v$. The lateral neurons have been shown to be both necessary and sufficient for flies to exhibit rhythmic eclosion and locomotor activity; flies expressing *per* only in the LN clusters remain rhythmic (Frisch *et al.*, 1994) while *disco* mutants

which lack functional LNs are arrhythmic in eclosion and locomotor assays (Ewer *et al.*, 1992). Most strikingly, the rare presence of a single functional LN in a *disco* fly is enough to restore rhythmicity (Helfrich-Förster, 1998). Molecular oscillations in the l-LN$_v$, however, which does not project into the dorsal protocerebrum, quickly dampen in constant darkness, suggesting primacy for the other two LN clusters as the source of the clock (Shafer *et al.*, 2002).

1.6.3 Seeking a pacemaker

"A crucial but still unanswered question in crustacean chronobiology is whether circadian rhythmicity is controlled by a single pacemaker or master clock located in the brain, or alternatively by several oscillators of (relative) independence... Unfortunately, mutant and genetic analyses comparable with those in Drosophila have only rarely ... or not (yet) been possible in crustaceans." - **Strauss and Dircksen (2010)**

The picture is much less clear as we turn to the crustacean clock. As the preceding quote highlights, the progress of research into crustacean chronobiology has been achieved with little recourse to the use of transgenics, reporter genes or binary expression systems. Instead, the typical approach has been to ablate or isolate tissues or structures considered as putative pacemakers and monitor the effects on rhythmic phenomena previously identified as under endogenous control.

In this manner evidence has mounted that, rather than possessing a central pacemaker acting as a 'master' clock, as seen in *Drosophila*, the rhythmic output of crustaceans – or, at least, *Procambarus* crayfish, the subject of much of the work – is governed by a complex system of oscillators. Thus it is seen that electroretinogram (ERG) rhythmicity persists in the isolated retinal tissue of the crayfish (Arechiga and Rodriguez-Sosa, 1998); the effect of eyestalk extracts on the amplitude of the ERG response in isolated eyestalk tissue depends on the circadian time at which the extracts were prepared (Moreno-Sáenz *et al.*, 1987); yet locomotor rhythms persist, albeit transformed, in crayfish with ablated eyestalks and lesion of circumesophageal connectives to the thoracic ganglion suppresses rhythmic locomotor activity in certain walking legs, implicating the supraesophageal ganglion as a source of rhythmicity (Page and Larimer, 1975); brain resection fails to suppress locomotor rhythms (Fuentes-Pardo and Rubio, 1981); the isolated caudal photoreceptor (CPR), a light-sensitive interneuron in the 6th

abdominal ganglion, maintains a circadian rhythm of electrical activity in constant darkness (Rodríguez-Sosa *et al.*, 2008); and ganglionectomy of the CPR affects ERG response and locomotor rhythms (Fuentes-Pardo and Inclán-Rubio, 1987).

The model of the crayfish clock built up from this work is one of multiple pacemakers required for the generation and maintenance of full rhythmic function, as shown in the schematic in Figure 1.5. This multi-oscillator model also fits with evidence showing the existence of multiple circatidal systems in the shore crab *Carcinus maenas*, which entrained to three different tidal *zeitgebers* presented out of phase with each other, subsequently showing three peaks of locomotor activity corresponding to the 'high tide' of each (Warman and Naylor, 1995). But ablation and lesions are crude interventions, sledgehammers compared to the array of scalpels on offer to researchers of *Drosophila*. More fine-grained investigation into the crustacean clock requires a model organism of comparable power.
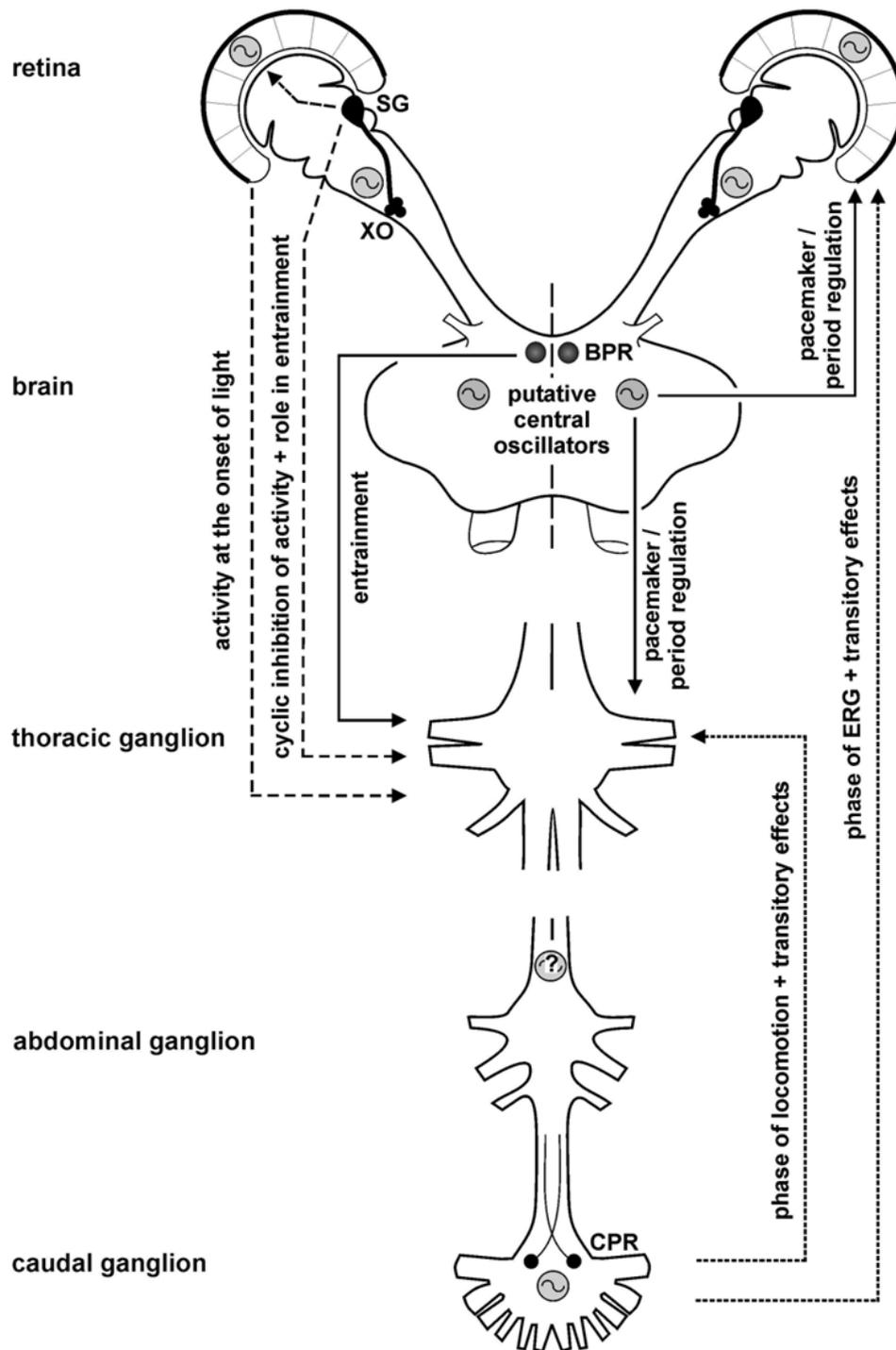
Figure 1.5: The oscillators of the crayfish circadian clock. Pacemakers (grey circles with ~) have been identified in the retina, the X-organ (XO) sinus gland (SG) complex within the eyestalk, the brain and the caudal photo receptor (CPR). Brain photoreceptors (BPR) are involved in light-mediated entrainment but are not believed to possess intrinsic circadian oscillator ability. The ? in the abdominal ganglion indicates a hypothesised oscillator in the ventral nerve cord. From Strauss and Dircksen (2010).

1.6.4 *Parhyale hawaiensis*: a model organism *par excellence*?

Over the past decade, the amphipod *Parhyale hawaiensis* (hereafter *Parhyale* unless otherwise stated) has become a model organism of some significance. It was originally collected for lab culture from the filtration system of the John G. Shedd Aquarium in Chicago in 1997, making it, as the collectors themselves put it, "preselected [for] minimal care"(Browne *et al.*, 2005). In nature it is a circumtropical marine detritivore found in intertidal, shallow water mangrove zones. The frequent changes in salinity, temperature and turbidity that characterise such a habitat have produced a particularly robust species well-suited to lab culture; like *Drosophila*, it is fecund and cheap to maintain. It breeds year round, with a quick generation time and a reproductive system that allows the collection, study and manipulation of embryos.

This fundamental suitability to lab research has seen *Parhyale* feature in a number of developmental studies since the first cultures were established in 1997 (Gerberding *et al.*, 2002; Liubicich *et al.*, 2009; Nestorov *et al.*, 2013), and also work exploiting its regenerative capabilities (Konstantinides and Averof, 2014). During this period it has proved itself to be similarly amenable to molecular manipulation and techniques successfully established for the species include *in situ* hybridisation, RNA interference mediated gene knockdown, transposon and integrase-based genetic transformation (Kontarakis and Pavlopoulos, 2014) and, very recently, CRISPR/Cas9 mutagenesis (Martin *et al.*, 2015). The genome is estimated at 3.6 Gb (Parchem *et al.*, 2010) and has recently been sequenced with a draft genome assembled (Kao *et al*, unpublished).

No research has yet been conducted on rhythmic phenomena in *Parhyale*, save a single study of feeding preferences that did not address the question of endogenous control (Poovachiranon *et al.*, 1986). Any findings in this area may be broadly applicable to other malacostracans such as *Euphausia superba*, *Eurydice pulchra* and the various crabs, lobsters and crayfish discussed above, as well as informing the view of the evolution of the circadian clock.

1.7 Research aims and outline

This project was undertaken to expand our knowledge and understanding of the circadian clock in *Euphausia superba* and Crustacea and to contribute to the study of

the evolution of the clock. In addressing the broader questions therein a second thread developed with the aim of investigating the suitability of *Parhyale hawaiensis* to circadian research, with the view to establishing this promising species as a model organism in crustacean chronobiology.

Therefore Chapter Two covers the cloning and characterisation of the canonical circadian genes in *Euphausia*, including sequence data, phylogenetic analysis, gene expression and protein interactions. In Chapter Three, details are given on the assembly and annotation of the *Euphausia superba* Transcriptome Database, a resource built and mined for information of genes related to the regulation and output of the clock. Chapter Four covers preliminary findings on the rhythmic nature of locomotor activity in *Parhyale.* Finally, Chapter Five describes the assembly, annotation and mining of a transcriptome for *Parhyale hawaiensis*, as well as data from the new draft genome.

Chapter 2   The cloning and characterisation of the canonical clock genes of *Euphausia superba*

2.1   Introduction

2.1.1   The core genes of the circadian clock

Chapter One contains an overview of the molecular underpinnings of the circadian clock in *Drosophila melanogaster* and *Mus musculus* (see 1.2). In the following sections the key components and their interactions are discussed in more detail.

2.1.1.1 The central heterodimer

The gene *Clock* was identified and characterised in *Drosophila* by Allada *et al.* (1998) through the forward genetics approach, using chemical mutagenesis to screen flies with aberrant locomotor activity rhythms – one of those identified was initially dubbed *Jrk*. Around 50% of heterozygous *Jrk* mutants were arrhythmic in constant conditions while the rest exhibited a longer period of activity compared to wild type; homozygotes, meanwhile, were totally arrhythmic in both locomotor behaviour and eclosion assays. The study also identified low levels of PER and TIM proteins and their respective mRNA transcripts, implicating *Jrk* in their synthesis. The mutated gene was identified as a homolog of the mouse *Clock* gene (*mClock*), itself identified through generation of arrhythmic mutants (Vitaterna *et al.*, 1994) and the name assigned accordingly.

Drosophila Clock (*dClock*) encodes a basic helix-loop-helix (bHLH) PAS protein with a notable C-terminus feature – an extensive polyglutamine (poly-Q) section in which 47 of 57 sequential amino acids are glutamine. Such a trait is common to *Clock* genes across species (Saleem *et al.*, 2001), though highly variable in extent.

Cycle (Rutila *et al.*, 1998) similarly encodes a bHLH-PAS protein, around half the size of dCLK. Homozygous mutant flies are arrhythmic, while heterozygotes have altered periodicity, and again *per* and *tim* transcription and protein abundance is affected. *Drosophila* CYCLE (dCYC) is something of an anomaly, in that it entirely lacks a C-terminus domain, the BMAL1 C-Terminal Region (BCTR), that is seen in the mouse

homolog BMAL1 (Gekakis *et al.*, 1998) and is present and highly conserved across many other species (Table 2.1).

Table 2.1: Conservation of BMAL1 C-Terminal Region across species. Orange – Crustacea. Blue – Insecta. Yellow – Chordata. See Appendix I for accession numbers.

Organism	Alignment of C-terminus
Daphnia pulex	QSSSRENSSP-GD---GNDEAAMAVVMSLLEADAGLGGPVDFSGLPWPLP
Calanus finmarchicus	NSPPSSSASS--E---GNDDAAMAVIMSLLEADAGLGGPVDFSGLPWPLP
Eurydice pulchra	---VESEVTS------DSDEAAMAVIMSLLEADAGLGEPVDFSHLPWPLP
Pacifastacus leniusculus	DVVSGRDLET--DGTSDSDEAAMAVIMSLLEADAGLGGPVDFSHLPWPLP
Tigriopus californicus	SSEMDRGPKE--G---GNDEAAMAVIMSLLEADAGLGGPVDFSGLPWPLP
Gryllus bimaculatus	PDSPHPSAFN--D---GNDEAAMAVIMSLLEADAGLGGPVDFSGLPWPLP
Acyrthosiphon pisum	TSSPQQLSNL--G---TNDEAAMAVIMSLLEADAGLGGPVDFSGLPWPLP
Danaus plexippus	VSPPLPPLGL--D---GNGEAAMAVIMSLLEADAGLGGPVNISGLPWPLP
Antheraea pernyi	VSPPLPSLGI--D---GNGEAAMAVIMSLLEADAGLGGQVNFSGLPWPLP
Lutzomyia longipalpis	TEAPNQTGTT--D---GNDEAAMAVIMSLLEADAGLGGPVDFTGLPWPLP
Aedes aegypti	EVPHSQANST--D---GNDEAAMAVIMSLLEADAGLGGPVDFSGLPWPSP
Anopheles gambiae	DVSQTQASST--D---GNDEAAMAVIMSLLEADAGLGGPVDFSGLPWPLP
Thermobia domestica	PESPNQNVSN--D---GNDEAAMAVIMSLLEADAGLGGPVDFSGLPWPLP
Tribolium castaneum	ESPSNEPVSE------GNDEAAMAVIMSLLEADAGLGGPVDFSGLPWPLP
Athalia rosaea	NESPNPMPTD------GNDEAAMAVIMSLLEADAGLGGPVDFSGLPWPLP
Apis mellifera	NESPNPVPSD------GNDEAAMAVIMSLLEADAGLGGPVDFSGLPWPLP
Nasonia vitripennis	NESPNPVPTD------GNDEAAMAVIMSLLEADAGLGGPVDFSGLPWPLP
Danio rerio	----EPGSSS--P---SNDEAAMAVIMSLLEADAGLGGPVDFSDLPWPL-
Gallus gallus	----DQGSSS--P---SNDEAAMAVIMSLLEADAGLGGPVDFSDLPWPL-
Ovis aries	----DQGSSS--P---SNDEAAMAVIMSLLEADAGLGGPVDFSDLPWPL-
Mus musculus	----DQGSSS--P---SNDEAAMAVIMSLLEADAGLGGPVDFSDLPWPL-
Xenopus laevis	----EHGSSS--P---SNDEAAMAVIMSLLEADAGLGGPVDFSDLPWPL-

A bHLH domain allows DNA binding to E-box sequences (Jones, 2004) while PAS domains enable protein dimerisation (Huang *et al.*, 1993), and so it goes with dCLK and dCYC, which form a heterodimer to drive transcription of clock controlled genes (CCGs) with upstream E-boxes (Darlington *et al.*, 1998). The same heterodimer formation and activity is seen in the mouse clock (Hogenesch *et al.*, 1998; Gekakis *et al.*, 1998), but there is a notable variation in how the work gets done. In the fly, transcriptional activation (transactivation) activity appears to depend on the extensive, glutamine-rich C-terminus of dCLK, deletion of which abolishes the heterodimer's activity and underlies the arrhythmic behaviour of *Jrk* mutants. In mice, the effect of varying C-terminus deletions of mCLK range from no effect to a 50% reduction in transactivation activity, and the protein is thought to lend only support to enable the activity of the other half of the dimer; the vital component is the BCTR of mBMAL1, deletion of which

abolishes activity (Takahata *et al.*, 2000). As dCYC lacks the BCTR, it seems that in *Drosophila* dCLK has adopted the role of main transactivation component to compensate for this apparent loss.

2.1.1.2 Repressive elements

Period was the first circadian gene to be identified in any species, achieved through the analysis of mutant *Drosophila* showing altered rhythms in locomotor activity and eclosion pattern (Konopka and Benzer, 1971) and even male courtship song (Kyriacou and Hall, 1980). The gene encodes a PAS protein – indeed the PAS (Per-Arnt-Sim) domain is partly named for it – but, in contrast to CLK and CYC, it lacks a bHLH domain, suggesting it functions through protein interactions rather than DNA binding. Upstream of *period* lies a stretch of DNA containing an E-box sequence implicated in the gene's rhythmic transcription (Hao *et al.*, 1997). In mice, three PERIOD (mPER) paralogs contribute to the clock (Jin *et al.*, 1999).

Mutagenesis through the use of a transposable P-element identified the *Drosophila* clock mutant *timeless* (Sehgal *et al.*, 1994), mutation once again impacting locomotor activity and eclosion rhythms and implicated in interactions with PER through its effect on *per* transcript cycling. The mutated sequence was identified to be a deletion that truncates TIMELESS (dTIM) to just over half the length of the wild type peptide (Myers *et al.*, 1995), which was surprisingly found to show no homology to dPER despite evidence that the two physically interact at the site of dPER's PAS domain (Gekakis *et al.*, 1995). Again, the *timeless* promoter contains E-box sequences so that expression of the gene is driven by the CLK:CYC heterodimer (McDonald *et al.*, 2001).

Embryonic *Drosophila* S2 cells assays have shown that when expressed separately dPER and dTIM accumulate in the cytoplasm, mediated by cytoplasmic localisation domains, while when coexpressed they move as a heterodimer to the nucleus 4 hours after heat-shock induction of expression (Saez and Young, 1996). The heterodimer, in which dTIM stabilises dPER to prevent degradation (Price *et al.*, 1995) has also been confirmed *in vivo* (Zeng *et al.*, 1996). In the nucleus, the components of the dimer interfere with the transactivation activity of the CLK:CYC heterodimer through inhibition of its DNA binding activity (Lee *et al.*, 1999) and thus repress their own

expression (Darlington *et al.*, 1998), closing the negative feedback loop that forms the basis of the clock.

In the mouse circadian system, however, the gene named *Timeless* contributes little to circadian timekeeping. Instead, the mammalian negative feedback loop is completed through the nuclear translocation and activity of complexes formed from mPER and the protein products of cryptochromes (Kume *et al.*, 1999). Mammalian *Timeless* is in fact the mammalian ortholog of the *Drosophila timeless* paralog *timeout* (Benna *et al.*, 2000)*,* and plays a crucial early developmental role (Gotter *et al.*, 2000). In *Drosophila timeout* loss is similarly lethal, and has been linked with circadian photoreception (Benna *et al.*, 2010). Perhaps unsurprisingly, given its importance, the gene appears to be ubiquitous in animals (Sandrelli *et al.*, 2008).

2.1.1.3 The cryptochromes

Cryptochromes are photoreceptor flavoproteins, first described in plants and derived from blue-light sensitive DNA (6-4) photolyases (Devlin and Kay, 1999), and shown to play vital – yet different – roles in both the fly and mouse circadian clocks. In the former, it is the light-input component of the clock.

On exposure to light dTIM is degraded (Myers *et al.*, 1996), allowing in turn the degradation of dPER and the release of their repressive effect on the dCLK:dCYC heterodimer. The ubiquitin-proteasome mediated degradation of dTIM (Naidoo *et al.*, 1999) is promoted by the binding of dCRY, the product of *Drosophila cryptochrome* (Emery *et al.*, 1998), in a light-dependent manner, with binding *in vivo* inhibited in the dark (Busza *et al.*, 2004). The light-dependency of this action has been isolated to the C-terminus with the generation of CRYΔ (Rosato *et al.*, 2001), in which the deletion of 20 residues allowed the protein to bind to TIM in both light and darkness. Further evidence of the importance of dCRY as a photoreceptor is seen in a study in which mutant cry^b flies maintained rhythmicity in constant light (Emery *et al.*, 2000) in contrast to wild type flies which become arrhythmic (Konopka *et al.*, 1989) – the phenotype, in which the fly's clock appears 'blind' to the light, can be rescued by expression of wild type dCRY. It is important to note, however, that dCRY is not the

only photoreceptor in *Drosophila* – the compound eyes, ocelli and extra-retinal Hofbauer-Buchner eyelets are also implicated (Helfrich-Förster *et al.*, 2001).

In mice, as noted above, the cryptochromes play a different role, interacting with mPER to form the repressive part of the feedback loop. Rhythmicity is lost in *mCry1* and *mCry2* mutants (van der Horst *et al.*, 1999) and expression of *mCry1* and *mCry2* is reduced in homozygous *Clock* mutants, suggesting they are driven by the central mCLK:mBMAL1 heterodimer (Kume *et al.*, 1999). The latter study also showed, through a luciferase assay, that both proteins enter the nucleus and are capable of inhibiting transcription by more than 90%, proving far more potent repressors than mPER alone. It has been determined that while mPER associates with the N-terminal half of mBMAL1, deletion of the BCTR of the latter prevents association of mCRY1, indicating that the BCTR is the crucial domain of both expression and repression in mice (Kiyohara *et al.*, 2006).

Neither mCRY is light sensitive, with mPER1 instead implicated in the light entrainment pathway in mammals (Field *et al.*, 2000).

2.1.1.4 Tying it all together

To summarise: in *Drosophila* the dCLK:dCYC heterodimer drives the expression of CCGs through DNA binding to E-box sequences, with transactivation activity dependent on the glutamine-rich C-terminus of dCLK. Two of these CCGs, dPER and dTIM, dimerise and enter the nucleus to repress the activity of dCLK:dCYC. Light-activated dCRY targets dTIM for degradation, releasing the repressive effect and allowing the cycle to begin again. In *Mus musculus*, the mCLK:mBMAL1 heterodimer depends on the latter component's BCTR for transactivation activity, and the repressive limb of the loop is composed of mPER:mCRY complexes, with mPER1 enabling light input.

A highly conserved system, but one in which some components vary in their function – an unsurprising finding given the hundreds of millions of years since the two species diverged (Ayala *et al.*, 1998). But what of other animals - and what can they tell us about the evolution of the biological clock?

2.1.2   Other organisms

While clock genes have been identified in many species, in depth analyses of their function are understandably rather fewer, but those that have been conducted – often using cell culture expression systems – provide a fascinating insight into how the clock may have evolved.

2.1.2.1 Insecta

Chang *et al.* (2003) used *Drosophila* S2 cells to characterise the clock gene activity of the silkmoth *Antheraea pernyi*. The results showed, as might be expected, that apCLK and apBMAL dimerised and proved capable of driving the expression through binding to E-box sequences, and apPER was capable of repressing this activity, an effect augmented with low doses of apTIM. But the *A. pernyi* heterodimer resembles the mammalian system more than *Drosophila*'s in that apBMAL possesses a BCTR crucial for transactivation activity, while truncation of apCLK has no effect (and it should be noted that apCLK also lacks a poly-Q in the C-terminus). Furthermore, the moth was found to have two cryptochromes – a *Drosophila*-like apCRY1 and a mammalian-like apCRY2 (Zhu *et al.*, 2005), with the former showing light sensitivity and the latter being light-insensitive but capable of repressing the activity of the CLK:CYC heterodimer of another lepidopteran, the monarch butterfly *Danaus plexippus* (Yuan *et al.*, 2007), itself a species that has been studied in detail (Zhu *et al.*, 2008). In *D. plexippus*, it has been demonstrated *in vivo* that dpTIM is degraded through the activity of light-activated dpCRY1 as with *Drosophila*, but the repressive limb of the loop is again a mammalian-like dpCRY2, which is degraded after the intervention of the dpCRY1 cascade. Neither dpPER nor dpTIM show a repressive effect and the authors suggest that the "CRY-centric" clock of *Danaus plexippus* may be an ancestral invertebrate type – the mosquitoes *Anopheles gambiae* (Zhu *et al.*, 2005)and *Aedes aegypti* (Yuan *et al.*, 2007) also possess both CRYs.

Drosophila is not the only insect with only one cryptochrome, however. The red flour beetle *Tribolium castaneum* and the honeybee *Apis mellifera* both possess only a mammalian-like CRY, and the latter also lacks a 'true' TIMELESS, with only a TIMEOUT confirmed present in the genome (Rubin *et al.*, 2006). Interestingly, this particular system – a mammalian CRY and no TIMELESS – holds true for all hymenopterans so far

studied, including the ant *Solenopsis invicta* (Ingram *et al.*, 2012) and the wasp *Nasonia vitripennis* (Werren *et al.*, 2010).

2.1.2.2 Crustacea

Veterans of Chapter One will perhaps recall the lament regarding the sequence-only basis of a great deal of crustacean clock gene data – much of it has been derived from assembled genomes or transcriptomes, with little subsequent functional or expression analysis. Zhang *et al.* (2013) is an exception, an in-depth study of the functioning of the circadian and circa-tidal clock of the isopod *Eurydice pulchra* that both identified the molecular components of the clock and employed S2 cell luciferase reporter assays to examine their workings. Of the canonical clock genes described above, *Eurydice* appears to lack only a *Drosophila*-like *cryptochrome.* In the luciferase assay, EpCLK and EpBMAL1 dimerise and drive expression through E-box binding, and a modest repressive effect is seen with cotransfection of either *Epper* or *Eptim* individually – expression is reduced to 70 - 80% when using an excess of repressor (100 ng *Epper*, 1 ng *Epclk*, 1 ng *Epbmal1,* for example). EsCRY2 proves a much more potent repressor, with expression reduced to around 10% with cotransfection of 50 ng of *Epcry2*. The *Eurydice* clock, then, appears similar to that of Lepidoptera, aside from the lack of a light-sensitive CRY. As the genome of this animal is not yet sequenced, however, such a component should currently be filed under 'not found' rather than 'does not exist'.

2.1.3   Research aims

Two *cryptochrome* genes have been identified in *Euphausia superba* (Mazzotta *et al.*, 2010; Özkaya *et al*, unpublished). This chapter covers efforts to clone and characterise the krill's remaining canonical clock genes*,* encompassing peptide sequence analysis, expression data and the investigation of protein interactions. The ultimate aim is to build a model of the Antarctic krill's circadian system and interpret the results in an evolutionary context.

2.2 Methods

2.2.1 Sampling

2.2.1.1 *Euphausia superba* capture and storage

Collection of research specimens took place during Antarctic summer of 2008, on the Discovery 2010 cruise JR177. *Euphausia superba* were identified using an EK60 echo sounder at around 60°S near the South Orkney Islands and North-West of South Georgia at 52°S. Animals were caught by target fishing using the RMT8 pelagic net. Krill to be used for identification of canonical clock genes were either taken from each catch and stored in RNAlater (overnight at 3°C and then at -20°C) or, for high volume catches, up to 200 krill were taken and frozen in a methanol bath at -80°C and then moved to a -80°C freezer for long term storage. From here on these samples will be referred to as **catches.**

2.2.1.2 RNA extraction and cDNA synthesis

Total RNA for use in degenerate PCR was extracted from catch samples, 4 krill heads taken from time points 1am, 6am, 1pm and 8pm. Total RNA was extracted using TRIzol reagent (Invitrogen) and stored in isopropanol and 3M NaAc. For cDNA synthesis RNA was precipitated from each timepoint and quantified using Nanodrop; aliquots from each timepoint were then combined and resuspended in a single mixture using RNase free water. Separate cDNA templates using either Oligo(dT) or Random primers were synthesised from 1 μg RNA using ImProm-II Reverse Transcriptase (Promega) and stored at -20°C when not in use.

For the analysis of clock gene expression in specific body parts, total RNA was extracted and cDNA synthesised as above from body parts dissected from samples stored in RNAlater. Body parts used were the antennae, eye, head, thorax and abdomen.

First strand cDNA for use in 5' and 3' RACE extension was created using RNA extracted from catch samples per the degenerate PCR extraction. RNA was precipitated from each time point and combined in RNase free water. Synthesis of RACE-ready cDNA was performed using the Superscript III RT Module from the GeneRacer kit (Invitrogen).

2.2.2 Gene cloning

In this chapter and those that follow, the results of many PCR cloning procedures are related. The specific polymerase will be noted in the Methods in each case, as well as any deviations from manufacturers' instructions. More generally, PCR products generated for sequencing were run on agarose gels (0.5 – 1.5% SeaKem LE agarose, 50 mL total volume; 3 µl EtBr or 5 µl GelGreen (Biotium Inc)) and visualised with either the Gene Genius (Syngene) or Dark Reader Transilluminator (Clare Chemical). Bands were cut from the gel and extracted using QIAquick Gel Extraction kit (Qiagen) and cloned into either pCR®4-TOPO® plasmid using the TOPO TA Cloning Kit for Sequencing (Invitrogen), or into the pMiniT vector using the NEB® PCR Cloning Kit (NEB). Electrocompetent DH5α cells or chemically competent 10β cells were transformed, plated onto Luria Bertani (LB) agar plates with the appropriate selective antibiotic and grown at 37°C overnight. Single colonies were selected for overnight culture at 37°C in liquid LB with selective antibiotic, shaken at 250 rpm, and plasmids purified using E.Z.N.A. Plasmid DNA Mini Kit (Omega). Recombinant plasmids were confirmed with a restriction digest using NEB enzymes and relevant buffers, allowed to proceed for 1 hour and then run on an EtBr gel and visualised. Sanger sequencing was performed by GATC Biotech using the ABI 3730xl DNA Analyzer. All the above protocols were performed per manufacturers' instructions, with the exception of an extended final spin of 2 minutes at > 13,000 g prior to elution during the gel extraction stage to ensure removal of excess alcohol and improve ligation success. Primers were designed manually or using Primer 3 (http://bioinfo.ut.ee/primer3/).

2.2.2.1 Degenerate PCR

Known protein sequences in other species were retrieved from the NCBI protein database (http://www.ncbi.nlm.nih.gov/protein/) using the recognised names/synonyms for CLOCK and CYCLE/BMAL1/ARNTL (Table 2.2) and aligned using Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/). Conserved regions were identified within these alignments and degenerate primers designed on this basis (Table 2.3). Degenerate PCR was performed using KAPA Taq polymerase according to the manufacturer's instructions.

Table 2.2: Sequences used to identify conserved regions of CLOCK and BMAL1/CYCLE.

Protein	Species	Accession
CLOCK	*Mus musculus*	AAC53200
	Gallus gallus	NP_989505
	Xenopus laevis	NP_001083854
	Danio rerio	NP_571032
	Drosophila melanogaster	NP_523964
	Thermobia domestica	BAJ16353
	Antheraea pernyi	AAR14936
	Macrobrachium rosenbergii	AAX44045
CYCLE/BMAL1/ARNTL	*Mus musculus*	NP_031515
	Gallus gallus	AAL98706
	Xenopus laevis	AAW80970
	Danio rerio	AAF64394
	Drosophila melanogaster	NP_524168
	Antheraea pernyi	AAR14937
	Thermobia domestica	BAJ16354

PCR products were identified through gel analysis, extracted and plasmid cloned, and recombinant plasmids purified and sequenced. Sequences were translated and the protein sequence subject to a reverse-BLAST search against the online NCBI non-redundant (NR) database to confirm identity. On confirmation, the fragment was reamplified using specific primers (Table 2.3) and KAPA Taq polymerase.

2.2.2.2 Searching the SRA

From the Sequence Read Archive Nucleotide BLAST page (NCBI), the SRA experiment set was restricted to SRX026165, the product of 454 pyrosequencing and at the time the only available transcriptomic data for *Euphausia superba* (Clark *et al.*, 2011). This database was queried with blastn using the coding sequences of *period* and *timeless* from *Drosophila* (accessions NP_525056 and NP_722914) *Daphnia pulex* (EFX76293 and EFX87311) and *Gryllus bimaculatus* (BAJ16356). For any hits with an alignment score of greater than 50 the contig was retrieved from the archive, stripped of adaptor sequences and translated, and the peptide sequence used in a blastp search of the NCBI NR database to confirm identity. The nucleotide sequence of the contig was also used to requery against the SRX026165 database to identify overlapping contigs that might extend promising sequences.

Specific primers (Table 2.3) were designed to confirm by KAPA Taq PCR contigs that were identified as fragments of *Euphausia period* and *timeless*. Once confirmed, these fragments were used to design RACE primers for extension.

2.2.2.3 RACE extension

RACE extension was performed using Platinum Taq DNA Polymerase High Fidelity (Invitrogen) according to GeneRacer kit instructions, with the exception that equal 1 μL volumes of GeneRacer primers and gene-specific primers (Table 2.3) were used. Each phase of RACE extension consisted of an initial PCR reaction followed by a nested PCR performed on the initial reaction product. PCR products from nested reactions were extracted, cloned, sequenced and analysed. Where apparently incomplete RACE products were obtained, further gene-specific RACE primers were designed and the process repeated in further phases until the sequence was believed to be complete.

2.2.2.4 Transcriptome mining

An early version of the *Euphausia superba* Transcriptome Database (Est-DB; see Chapter Three) was assembled while RACE extension of *Esperiod* and *Estimeless* was ongoing. Once it was available as a resource the sequences constructed from RACE up to that point, as well as the orthologs of other species, were used to query the database to attempt to identify the full coding sequences. Primers were designed on the basis of the results in order to confirm the full coding sequences by PCR.

2.2.2.5 Confirmation

All putative complete coding sequences were confirmed by PCR using Q5 High Fidelity DNA Polymerase (NEB) and the primers shown in Table 2.3, using an annealing temperature determined through use of the NEB Tm Calculator (http://tmcalculator.neb.com/#!/).

Table 2.3: Primers used in cloning the core circadian genes of *Euphausia superba*. See Appendix III for thermal profile details. Profile 24 is a standard Q5 polymerase profile used for a number of reactions with only the annealing temperature changed, hence that detail is provided here.

Gene	Use	Forward	Reverse	Thermal profile	Annealing temp (°C)
Esbmal1	Degenerate PCR	GICGDMGIGAYAARATGAA	TTYIRCRTCDATIARIAKYTT	1	
	Degenerate PCR nested	TIACBGTIYTIMGVATGGC	GAIARYTGYTCYTTIACYTT	2	
	Semi-degenerate PCR	TIACBGTIYTIMGVATGGC	GGATGAAGGATGTCAAACCATGATGTCC	15	
	Fragment confirmation	ATGGCTGTACAACACATGAAGATGC	GGATGAAGGATGTCAAACCATGATGTCC	15	
	Tissue specific PCR	ATGGCTGTACAACACATGAAGATGC	GGATGAAGGATGTCAAACCATGATGTCC	15	
	RACE 5' phase 1		TCCATCGGCAGCCTGCAATATCAAG	3	
	RACE 5' phase 1 nested		ACGCAGCACTGTGAGCTTGTCTAAC	5	
	RACE 3' phase 1	CCTACATTATTCACAGAGTGAGCTTCTT		4	
	RACE 3' phase 1 nested	GGACATCATGGTTTGACATCCTTCATC		6	
	RACE 3' phase 2	GCGACTGCAAAGTACCTGGCGAAC		11	
	RACE 3' phase 2 nested	CTGCAAAGTACCTGGCGAACATTCA		12	
	Full coding sequence	CATTCAATCATGTTCGGTCTGG	CAGTCTTCGCAGATTATGGCA	24	65
Esclock	Degenerate PCR	GARAAGAARMGWMGAGATCA	TCWGGYYTBGAATTSMACTG	1	
	Degenerate PCR nested	GARATWCRACAGGAYTGGAA	GTYTGNARCCADATCCA	2	
	Fragment confirmation	TTCCTCTCCAATGAGGAGTTCACG	GCGATAGTAACATGATGTACCCTTTCCAG	15	
	Tissue specific PCR	TTCCTCTCCAATGAGGAGTTCACG	GCGATAGTAACATGATGTACCCTTTCCAG	15	
	RACE 5' phase 1		GTGCGTGAACTCCTCATTGGAGAGG	7	
	RACE 5' phase 1 nested		TCATTGGAGAGGAAGGAGGGTTTCCA	9	
	RACE 3' phase 1	CTGGAAAGGGTACATCATGTTACTATC		8	
	RACE 3' phase 1 nested	GGTACATCATGTTACTATCGCTTTCT		10	
	RACE 3' phase 2	GGCTAACAGCAGCTCTCATAATCTTG		11	
	RACE 3' phase 2 nested	CAGCAGCTCTCATAATCTTGAAGACA		12	
	RACE 3' phase 3	GTAGGACCAAGCCTGACCATGATC		11	
	RACE 3' phase 3 nested	CAGAGAGTGAGAGTGACCGGGATCT		12	
	Full coding sequence	ATGCCATTACATGAATTCCTGAGG	GTGCATCTTTCACTTCTTGGGT	24	66
Esperiod	Fragment PCR	CCATATTTGGGTCATTTACCTCAG	CTTCTGGTCCCTTTATTACTCTATGT	17	
	Tissue specific PCR	AACAGAGGTGCAGCCATACC	TGCATTGATTGGCTTACCAG	18	
	RACE 5' phase 1		CAATAGTGGTAGATCTTCAGGATGGTAG	21	
	RACE 5' phase 1 nested		CACAGAGTTACCGGTGAGGTCCTGA	10	
	RACE 3' phase 1	TGGCTGTGCTTTGTGAATCCATGGAC		22	
	RACE 3' phase 1 nested	GGCCAACATAGAGTAATAAAGGGACCAG		23	
	RACE 3' phase 2	GGAACCTGCAGAGAAAAGACGACGT		22	
	RACE 3' phase 2 nested	AGACGACGTACCTTAGCACGAATGAT		23	
	RACE 3' phase 3	GTCCAGGATCCTACCGTCATGTTCA		22	
	RACE 3' phase 3 nested	CAGAAGCAGCGGACTCATAAACATTC		23	
	Full coding sequence	GGCTAAGGCTGGGCATAAC	ACTGTGACTTCAATGGGTTGTG	24	66
Estimeless	Fragment PCR	GAACCTCTTTGCTCAACGCTTTG	GAGACTGGATCCGACGTTGATATA	17	
	Tissue specific PCR	CCGCACAAAATGCTCAGTAA	GCTCTTCCTGCATCACAACA	20	
	RACE 5' phase 1		TCTCTTGCTTCTGCTCTCGGTACATC	21	
	RACE 5' phase 1 nested		CTCTCGGTACATCAAGGAAATCAGCTG	10	
	RACE 3' phase 1	TCCTTGATGTACCGAGAGCAGAAGC		22	
	RACE 3' phase 1 nested	GATGTACCGAGAGCAGAAGCAAGAG		23	
	RACE 3' phase 2	GAAGAAGTGCGATCATGCGGAGTCA		22	
	RACE 3' phase 2 nested	GGAGGCACTGAAGGAAGTTCTCTTC		23	
	Full coding sequence	GGTGGAAGAGCCGCTACATA	GGTATGACCTTGACTTCAGCT	24	65
Estimeout	Fragment PCR	CAAGGTGATGCAAAATGTC	CATCCTCATCTGCAATCTGG	24	63
	RACE 3' phase 1	CCAACACAGGAAAGGCTCTCTACTC		22	
	RACE 3' phase 1 nested	CAGATGAAGAGAATATTGACCCAGATG		23	
	RACE 3' phase 1 nested 2	GAGAATATTGACCCAGATGACAAGGA		23	

2.2.3 Gene characterisation

2.2.3.1 Tissue expression

PCR amplification was conducted on cDNA synthesised from antennal, eye, head, thorax and abdominal tissue using KAPA Taq polymerase and specific primers (Table 2.3) and the results run and visualised on an EtBr agarose gel.

2.2.3.2 Protein sequence analysis

Domains were identified using SMART (http://smart.embl-heidelberg.de/; (Schultz *et al.*, 1998; Letunic *et al.*, 2015), the NCBI Conserved Domain function and through the alignment of the peptide sequences with orthologous proteins and reference to specific literature. Instances of this latter approach will be supplied as necessary in the Results section.

Full peptide sequences, as well as domains of particular interest, were aligned with *D. melanogaster* and *M. musculus* orthologs using EMBOSS Needle Pairwise Sequence Alignment (http://www.ebi.ac.uk/Tools/psa/emboss_needle/) to ascertain sequence identity/similarity.

2.2.3.3 Phylogeny

Phylogenetic trees based on peptide sequences were generated using MEGA 7.0. Sequences (see Appendix I) were aligned with MUSCLE using default settings, and neighbour-joining trees generated using 1000 bootstrap replications and complete deletion. In the CLOCK tree some fragmentary sequences were used to give further representation to Crustacea, but only fragments that included the conserved bHLH and PAS domains were considered.

2.2.3.4 Transcriptional activity assay

The full coding sequences of *Esbmal1, Esclk, Esper, Estim* and *Escry2* were cloned into pAc5.1/V5-HisA (Invitrogen), an expression vector with a constitutive *D. melanogaster Actin 5C* (Ac5) promoter, using the Gibson Assembly Cloning Kit (NEB) per manufacturer's instructions. The primers for these reactions (Table 2.4) were designed using the NEBuilder Assembly Tool (http://nebuilder.neb.com/). Each insert was amplified from its original vector using Q5 High Fidelity Polymerase and the vector was

similarly linearised by PCR from a plasmid stock. Each recombinant plasmid was sequenced beginning upstream of the Ac5 promoter and through the entire insert to confirm that no polymerase errors had been incorporated. The vector's V5 and HisA epitope tags were kept in frame.

Table 2.4: Primers used in PCR for subsequent Gibson Assembly reaction, using thermal profile 25 with 2 minutes 30 seconds extension time.

Plasmid source	Forward	Reverse	Annealing temp (°C)
pAc5.1	TCTAGAGGGCCCTTCGAAG	TTTGGCCACTGTGCTGGATATC	64
Esbmal1	ATCCAGCACAGTGGCCAAAATGTTCGGTCTGGGGAATTATGAC	GAAGGGCCCTCTAGATGGCAGTGGCCATGGCAG	68
Esclock	ATCCAGCACAGTGGCCAAAATGCCATTACATGAATTCCTG	GAAGGGCCCTCTAGACTTCTTGGGTGGTAGTTG	60.3
Esperiod	ATCCAGCACAGTGGCCAAAATGACTGAAAATGAGAGATCTACTC	GAAGGGCCCTCTAGATGAGTCGGAGGAGTCTCTAC	62.4
Estimeless	ATCCAGCACAGTGGCCAAAATGGAGTGGATGATGATGAAC	GAAGGGCCCTCTAGAGTGTAAGTGAGCTGGCGC	62.7
Escry1	ATCCAGCACAGTGGCCAAAATGACCAACACTGGTTGTG	GAAGGGCCCTCTAGAACAGACATTGTGTTGACATTG	61.6
Escry2	ATCCAGCACAGTGGCCAAAATGACGAGAGGAGGAGAGAAACATG	GAAGGGCCCTCTAGAGGCCGGCTGCATTTGCTG	69.1

D. melanogaster embryonic S2 cells (Invitrogen) were initiated from frozen stock per manufacturer's instructions and maintained in complete medium (HyClone SFX-Insect Cell Culture Media, 10% fetal bovine serum, penicillin-streptomycin at 50 units penicillin G and 50µg streptomycin sulphate per mL of medium) at 25°C – all incubations during transfection took place at the same temperature. Cells were passaged 2 – 3 times a week by taking 2 mL of the current culture and adding it to 8 mL complete medium in a new 75 cm^2 flask.

Transfections were performed two days post-passage after counting and checking for 95 – 99% viability using trypan blue staining. Cells were diluted in complete medium to 5 x 10^5 cells/mL and 1 mL added to the wells of 12-well cell culture plates. As S2 cells are only semi-adherent plates were incubated overnight to encourage maximum attachment. Complete medium was carefully removed and replaced with 1 mL plating medium (HyClone SFX-Insect Cell Culture Media, 1.5% FBS). Cells were transfected using Cellfectin II Reagent (Thermo Fisher) per manufacturer's instructions using the plasmid constructs shown in Table 2.5 and incubated for 5 hours, after which the plating medium was carefully removed and replaced with 1 mL complete medium.

Transfected cells were incubated for 48 hours, then washed with PBS and lysed with passive lysis buffer (Promega). Luciferase activity was measured using Dual Luciferase

Reporter Assay Kit (Promega) and FLUOstar Omega (BMG Labtech) microplate reader, and E-box fused firefly luciferase activity normalised to *Renilla* luciferase activity.

Table 2.5: Transfection constructs and usage. Each transfection used 100 ng of E-box luciferase reporter plasmid and 100 ng of *Renilla* luciferase control, and the empty pAc5.1/V5-HisA plasmid was used to keep total DNA equal across all transfections. Each transfection (1 – 8) was repeated 3 times.

Construct	Detail	Transfection amounts (ng)							
		1	2	3	4	5	6	7	8
pAc5.1-Esbmal1	pAc5.1/V5-HisA with *Esbmal1* coding sequence insert	100	100	-	-	100	100	100	100
pAc5.1-Esclk	pAc5.1/V5-HisA with *Esclk* coding sequence insert	100	-	100	-	100	100	100	100
pAc5.1-Esper	pAc5.1/V5-HisA with *Esper* coding sequence insert	-	-	-	-	100	-	-	100
pAc5.1-Estim	pAc5.1/V5-HisA with *Estim* coding sequence insert	-	-	-	-	-	100	-	100
pAc5.1-Escry2	pAc5.1/V5-HisA with *Escry2* coding sequence insert	-	-	-	-	-	-	100	-
pGL3 *4E hs luc*	Firefly luciferase reporter with four tandem repeats of *dper* E-box	100	100	100	100	100	100	100	100
pCopia-*Renilla*	Control plasmid using *copia* promoter fused to *Renilla* luciferase	100	100	100	100	100	100	100	100
pAc5.1	Empty pAc5.1/V5-HisA plasmid	200	300	300	400	100	100	100	-

2.3 Results

2.3.1 Gene cloning

2.3.1.1 *Esbmal1*

Degenerate and subsequent semi-degenerate PCR produced a 340 base pair (bp) fragment that was found to encode a 112 amino acid (aa) peptide. Initial RACE extension obtained overlapping 5' and 3' fragments of 429 bp and 881 bp respectively, the latter of which was found to lack a stop codon, suggesting a possible mis-priming of the 3' GeneRacer primer. Further 3' extension produced a 1,170 bp fragment with a stop codon. PCR using primers covering the putative ORF of the constructed sequence produced a 2,017 bp coding sequence encoding a 664 amino acid peptide. A blastp search using this peptide sequence returned *bmal1a* in the crayfish *Pacifastacus leniusculus* as the top hit (accession AFV39705: E-value 0.0, identity 73%, coverage 93%).

2.3.1.2 *Esclock*

Degenerate PCR produced an 851 bp fragment that translated to a 283 aa peptide. RACE extension identified an overlapping 5' region of 736 bp and a 3' region of 1,380 bp that when translated lacked a stop codon. Further RACE extension of the 3' end produced a 2,255 bp sequence with a stop codon. PCR using primers covering the putative ORF of the constructed sequence produced a 4,032 bp coding sequence encoding a 1,344 aa peptide. A blastp search using this peptide sequence returned *clock* in the prawn *Macrobrachium rosenbergii* as the top hit (accession AAX44045: E-value 0.0, identity 79%, coverage 56%).

2.3.1.3 *Esperiod*

The search of the SRX026165 dataset retrieved reads SRR064845.118199.2 and SRR064845.827174.2. Primers based on these sequences produced a 265 bp fragment that translated to an 88 aa peptide. RACE extension identified overlapping 5' and 3' fragments of 1,480 bp and 577 bp respectively, the 3' fragment lacking a stop codon, Subsequent attempts at 3' RACE extension produced further fragments of 1,034 bp and 993 bp, and while both indicated a stop codon, blastp searches suggested this would be a truncated protein in comparison to orthologs. Searching the *Euphausia*

superba transcriptome database using the sequence identified thus far produced an apparent full coding sequence (contig ES.23_comp19428_seq0) extending beyond the RACE sequences. PCR using primers based on these sequences produced a 3,783 bp coding sequence encoding a 1,261 aa peptide. A blastp search using this peptide sequence returned *period-like protein* in the lobster *Nephrops norvegicus* as the top hit (accession ALC74274: E-value 0.0, identity 67%, coverage 81%)

2.3.1.4 *Estimeless*

The search of the SRX026165 dataset retrieved reads SRR064845.327774.2 and SRR064845.912501.2. Primers based on these sequences produced a 243 bp fragment that translated to an 80 aa peptide. RACE extension identified overlapping 5' and 3' fragments of 1,084 bp and 519 bp respectively. The 3' fragment lacked a stop codon and subsequent attempts at RACE extension were unproductive. Searching the *Euphausia superba* transcriptome database using the sequence identified thus far produced an amended 5' region (contig ES.k31.R5250249), and a search using the full TIMELESS sequence of *Eurydice pulchra* retrieved a large 3' fragment of 3415 bp including a stop codon (contig ES.23_comp16712_seq0). PCR using primers based on these sequences produced a 3,933 bp coding sequence encoding a 1,311 aa peptide. A blastp search using this peptide sequence returned *timeless-like protein* in the lobster *Nephrops norvegicus* as the top hit (accession ALC74275: E-value 0.0, identity 75%, coverage 46%).

Large fragments of a mammalian-style TIMELESS (paralogous to *D. melanogaster* TIMEOUT) were also identified during this process. Through fragment PCR and RACE extension the full 4,089 bp coding sequence encoding a 1,363 aa peptide was confirmed. A blastp search returned *timeout* in *Tribolium castaneum* as the top hit (accession EEZ99220: E-value 0.0, identity 51%, coverage 79%).

2.3.2 Protein sequence analysis

2.3.2.1 EsBMAL1

A basic helix-loop-helix domain, two PAS domains and a PAC domain were identified by SMART analysis (Figure 2.1). Also identified through reference to Chang *et al* (2003) was the BCTR common to BMAL1 proteins but not *D. melanogaster* CYCLE – hence the

former naming convention was adopted for this gene – and putative nuclear localisation and nuclear export signals through reference to Hirayama and Sassone-Corsi (2005).

```
MFGLGNYDFSGSYAGSDCASMGSFSESDVGSRKRRASCVDSNDEDGDSVKMSRNSGWSKRSFAGDQSESEGGKIS
TQHGRQNHSEIEKRRRDKMNTYIMELSSIIPVCTSRKLDKLTVLRMAVQHMKMLRGSLNSYTEGHYKPAFLSDDE
LKNLILQAADGFLFVVGCDRGRILYVSESVYETLHYSQSELLGTSWFDILHPKDLVKVKEQLSCSDIGRRERLVD
AKTLLPVKADVPQGLTRLCPGSRRAFFCRMRCKSAPVMKEEADSSTGCQKKKSKSQSSDRKYSVIHFTGYLKSWA
ETKDPLDEDSDSDSCNLSCLVAVGRVHQPLVSTSVRDGSNFNNTTHAQPIQFISKHTNDGKFVFIDQKATLLLGW
LPQELLGSSMYEYFHQDDIPALAETHRETLKTANSIDTLVFRFRTKDGSFVRLQSTWRTFKNPWTKDIEYIIAKN
SMVTSETGVVESTMANDSVTQSYGNFNDFISSPGDAEASGGSSRLLAAGSGVHAGKIGRQIADEVLDNSRRSDSP
AENPLSPFEGMLGSTVTDRSFAALLRTDLTTHRKPVKNNVLLSTSNTSSDCDSSKTSSTQPLTARILNARAANKH
HYHSHNIPQDGASTDEPGEEGLAGDGGGASDSDEAAMAVIMSLLEADAGLGGPVDFSHLPWPLP
```

Figure 2.1: Peptide sequence of *Euphausia superba* BMAL1. Turquoise = bHLH domain. Blue = nuclear localisation and export (white text) signal motifs. Red = PAS domain. Dark red/white text = PAC (motif C terminal to PAS) domain. Green = BCTR.

2.3.2.2 EsCLOCK

A basic helix-loop-helix domain, two PAS domains and a PAC domain were identified by SMART analysis (Figure 2.2). A poly-Q region has been defined by Chang *et al* (2003) as one in which 60% or more amino acids are glutamine. Following this definition with the additional criteria that the region must be 10 amino acids or longer – otherwise QQ(X) is a poly-Q and we will see them in overabundance – three poly-Q regions were identified, including one large region in which 116 out of 163 amino acids are glutamine (71%). Putative nuclear localisation and nuclear export signals were also identified.

```
MPLHEFLRLGDDAQSGYETSVMDDGEGDEKDDVKRKSRNLSEKKRRDQFNLLINELSVMVAANNRKMDKSTVLKA
TISFLKNQKEMSTRSQGQEVREDWKPSFLSNEEFTHLMLEALDGFIMTVSCSGRVLYTSESITPLLGHLPGDLSE
SHVYDLMLSEERNDMRRFLSNPALAPNPSINMDDTKEQYAISVHLKRSPTNISEEPSYERVKLTGYFERYTCPSE
DGVLDFSCSEAEDSMSVDRSMFINRQNTNLGGIFGDSPSQSGVQDTKLVFVAIGRMERPQLVREMMILEPTKTEF
TSRHSLEWKFLFLDHRAPTIIGYLPFEVLGTSGYDYYHVDDLDKVSNCHERLMKTGKGTSCYYRFLTKGQQWIWL
QTNYYITYHQWNSKPEFIVCTNTVVGYSAVKGQLVKEENGEDCVEDDEDDPLDLNRMYSAGPSGVRMGPMDADNR
SQCSGTGGNSSDDNRSLSSMAHSVGPSLTMIKSQPNQTESESDRDLLSSDSDGGQRRSIADILPSSQRQKQLQTL
SQRQQILDQQIQHQKRLQHEQEQRQHQPHQQQHQQPPQOPPQPPPQPLFSESDTEAIQNHQVQTLAEQQLQQVR
ALQQKHEEQQLQQAALHQQQEQLLQQQFVKPPPQPQPLQPPRPQPPQSLPPQQQQLHLQLPDPQVSPASPSSSQ
RQQPPHSPYPQPPHHLPSRGDPPPWSPKARPGSSKRLRIQQQQQQQQRQQQQQQHQQQSPQIYLDDMTSLPSPG
SDISVNSQTSHSSMQSHGSQHGSLQQTPDKRRLPQRDYQANSSSHNLEDTHDQVRLYIPKQLKYGSRTPLYKTGN
SDNETSPYTSKAKHKKYYRKIELPPEISSRLTYKKEEVEDQLMGSMVSSPGSCSESASVSGMSLGNVIVSSGQGV
VLSSPQHPQQQQQHHQQHIQQQEQQQRQQQQQQQQRQQEQRQQEQRQQEQQQEQQQRQQEQRQQEQRQQEQQQR
QQEQQQRQQEQQQRQQEQQQRQQEQQQRQQEQQQRQQEQRQQEQQQRQQEQQQRQQEQQQQQQQISGMY
GVVQQQQQQQQQQQQQQQQQYLGGALGGALGGGGSSSGSIRGHAAPSSSAAATQVLQVATVPFSPVLSIPGQPV
VANLSVQLPQPQTTEDLRPQLILSPEQRQLQEQLRLKHAELQQMIMRQQQELQTVNEQLVMAQYGMYKNNLQYAT
SPAIGSSSSGGFPGGVNVSGPGNMAGMSALQTVGGPGSGSLHASNNPGTATITLGNAATVSLGSSAAAQSNFIQTS
STPAPSSLQHNLQPTTITVTSINPRQVNVVSGGQRMAVPSIPLSHQQASMLFAQPVPGTAQPPQLPPKK
```

Figure 2.2: Peptide sequence of *Euphausia superba* CLOCK. Turquoise = bHLH domain. Blue = nuclear localisation and export (white text) signal motifs. Red = PAS domain. Dark red/white text = PAC (motif C terminal to PAS) domain. Yellow = poly-Q region (≥ 60% amino acids glutamine).

2.3.2.3 EsPERIOD

Two PAS domains, a PAC domain and a PERIOD-C domain were identified by SMART analysis (Figure 2.3). Alignment with *D. melanogaster* PERIOD suggested a region that may be a CLK:CYC inhibition domain (CCID; Chang and Reppert (2003)) although alignment of those regions alone showed only 22.1% identity and 33.2% similarity.

```
MTENERSTLVPTISVEESEQVPVEGHNVKSSRNPEPAPSQMDSAYGSLGSSSLSRSRGESCSGSKSRHSGSSQSS
ILGDHHAEFRGHAQIIKIKELKKPPVEKKITICEVKEEVKEKIIVEEPVFNVEFAPPKEKIEPEVNRTVITKPEE
NPNSNSQSQDVYQGISDQPTLVYTQALNYIRKIKERSAEQGLPFASPNLIQLPRDKQHGATLFKSLKSVRGFTVA
ISIQDGTALQVSPSITDVLGFPKDMMLGQSFIDFVYPKDSINFSSKIIHGLNMPFINESIKDSFGTSLFCRMRMY
HSLRNSGFGVRNKRTMYKPCKMILKFHDVSNIETDNPNYGQNSSMILAEVIPIDSVYKYPGEVPAMGCFSIRHSA
SCNFSDYDTEAIPYLGHLPQDLTGNSVFDCYHPEDLPLLKTIYEDIVMQQGRPHRSKPYRFRTFNGSYVTLETEW
LCFVNPWTMKIDSVIGQHRVIKGPEDIQIFLDPGERSYTKFSEEVLKEAQKAQNYIVELLAKPVASYKDPSKEPA
EKRRRTLARMMTSIVDELDHMEKKKSNIIQMPSTSKTMVVPAKIQRIVPKLPSTISKGQESLNSSSGTPSSYGQL
NYSETIHRFFRSIPKTISSDESGDSKNSPQGSLDDTKSSTQPEFSTSGASGSGNMKSQQSQQFSSCSGSGDTSDN
NRNKNPNSVTSNSPGSYRHVQLTEEIMNRHNVDQQHKFVQKQRTHKHSKDRFKIKNIKKATKRPMERTHGLKRPG
TIMDRECSAPKQIYMASSADDSGNGKCIKTGSSGVKSRDRPTSVHSINVFNSGGANRGAAIPSSSGLAPNFQHGM
MGHGMSGYYQIAAMPNFNLNGPSIENVPVASPNAVFMQPQQQQFLQPQSLMPGMAMQYMGPGMGAFPGIVYQTVG
PPYFNAPQLMLPNLMYQHTAMPQPQPVAQHIAAHMDSEEEKTILRTDSEHEIIDSQHITGNRIIISGKPINAHQR
PDSRATSVKAEPGSVRGSMARESSHSHAAESIRSYVEDVTHCFAPKESEKHPHSSQLSEFSQSSSVQAEAESVGS
PGKIDNKPNDELHASDIIMSVTSSTHESSKDLYYKSTDDSFLLNSESSEMSEGKFNNKWNTSKPKYQPPKLGGPS
WLENVDVTPELLYKYQLETKELVDVLRQDMDVLKKFRQPVCVDDQLSFLYNEMEREGESTKLHLEEGITSSSSGE
EGPCSSKDGAASTSRGRVKKSVYFSTAAMLQEEDAPMPPPESRESSYYLMCEEPGRDSSDS
```

Figure 2.3: Peptide sequence of *Euphausia superba* PERIOD. Red = PAS domain. Dark red/white text = PAC (motif C terminal to PAS) domain. Red text = putative CCID. Dark green = PERIOD C domain.

2.3.2.4 EsTIMELESS

SMART analysis identified an N-terminus TIMELESS domain, while the NCBI Conserved Domain Database (CDD) indicated that this was incomplete, the domain alignment beginning at residue 2 of the consensus sequence and ending at residue 213 of 266. The CDD further indicated the presence of a partial TIMELESS C domain (Figure 2.4), the alignment beginning at residue 148 of the consensus sequence and ending at residue 378 of 507. Alignment with *D. melanogaster* TIMELESS identified two putative PER interaction sites (PIS) as detailed by Ousley *et al.* (1998), showing 39.2% identity, 59.5% similarity (PIS-1) and 39.3% identity, 56.9% similarity (PIS-2) with the equivalent *Drosophila* sequences. The cytoplasmic localisation domain (CLD) described in the same paper was not detected.

SMART analysis of EsTIMEOUT identified TIMELESS and TIMELESS C domains (Figure 2.5), both of which NCBI CDD confirmed as complete with the alignments encompassing the full consensus sequence.

```
MEWMMMNMDRFYAPAVQLGNYYGDKYVPSKNYKAKIDEILRRLVQDKARWHFRRSLFMNKIISQNLLPTLTHVKE
DREVLEATVKVLQELMTPVECLIPMETMSRNCEGRRIIQELESSIMLHKKMFLDTRVTKAIVDLMGSTLQDSKKV
LSMAECECINHCLLLVRNLLHVSNNSLVQEQQQKQRQQNHNQKSEAGMNNGSASVLLEDQIMWNLFAQRFDNILI
QIFTSHQQHLWNITIVQLISLMYREQKQENIRNMINEWLDSSLSESSEDDENNTMSSSSDNISTSDPVSETSERL
LPVNEPIKLKINKQTSTSSSTINSEESQAITNKTQSAADSGFCSSFQPSGSNQDSNEETTIQVSCLHKQQSQITK
PADQGNEDKTDLNIEDLQFKDTAQKENQTNLVLAIEEEVRSCGVNEGGTEGSSLLHTLDDTTPLSCQRLEIEKAN
EIKEVTQIISMLEDGSKGETAANAPVSLEMAIEQTKLAIEQTKLAIEQTKLQQNTNLGKGGLNQLAQQDMSINNQ
FNMSACTIKTTPTDTNFTSPCVLQNLAQSELFRRPPNETEADSDDNKPPPQLPKLTKKNPGSGTKRSRHSMSQIM
SENQENEYSNSTGSDYDEGPACKRPHHQKPHKMLSKPRPAKMMQKAIQERNVKRTKLLRHKENYSIKAKALLHHH
PTSEDITNLLKEFTVDFMLKGYANLLEGLRLQVLMPYQIDLDKSHVLWLITYFIRFAVELDLELSILCPVLSVET
VSYLVYEGVVMQEELERATHNGEANLQPHIRRLHLVVTALREFFLAFEICMKKDPATYDSQQLLKIKEDLGQLVE
VRQLFVLLIRTYKPGVLNLNYLQDLITTNHRFLTTQEAASPTLSTLNTFDIFDHVKQFCTSEIMRQYGRLLENFE
NNDEQVNDCIFTMMHHTVGDLRSINCLMQPQILKIFLKIWNEGFDLCVDWSDLIEYVLRKCTIVRTESRKRQACK
NSSRRMGMMDNTKTKLTDEDLGMWYSLNYSVQENEAGEMRKIMEVGGDKSIEEPVKKEVIQKLLARGFITSIECTKL
CAEIPIIVERKSVTSVSLNSDHMQVSITDIETLSTGTQDPEDSLPAPKKCPEEVKDNPTRFRSMPEPIKDLAMCG
VDTDFESSQPFDREDPWDDNTNVIGFINKLKEEGLGAHVEWLQGQLLEACYAKLKILGPDLPRAEPIASHFTISN
QSIPLIPWSLDQDVVLSNPWFRQLLGALGLHQPNDTGKVFPRIPHFWTPDVLFLMAKRLGDIDSRSLTSLLWIKA
QHVWSSLITHSQAENRSACAPWKTIDQPTPAPAHLH
```

Figure 2.4: Peptide sequence of *Euphausia superba* TIMELESS. Black = TIMELESS domain. Red text = PER interaction site. Purple = TIMELESS C.

```
MSEGLLHAELQAACSALGYSDGAKYYKEADCLETVKDLIRYLKKDDESHEIRRALGETRVLQTDLIHLLRECSND
DELFDVTLRLIVNLTNPVLLLFHEELPEEKVTRNHFLQILSQQQTYKESFADEKLWSVLVGKLSELLQRDADTRQ
EEDGMIIERILIMIRNILSVPASPDEEKRTDDDASLHDQVLWALHLAGFEDLILYMASSSNEQNLCMHVLEIVSL
MLKEQDPEQLAQAKLQRSQVEKEKDEQELLETRRQEAIKKQQQFKKYHGARHSRFGGTFCVQNMKSISDRDIISH
KPLTDVQSINFDRTKRSKKIAKNRMPLLDKNLTRRSTLAIRLFLQEFCTEFLSGAYNFIMYIVKENLSRARGQEH
DESYYLWIMRFFLEFNRYHQFKVELVSETLHIQTFHYIQTNLETYFEMMTTDKKKIPLWSRRMHRGLGAYRELLM
TLALMDKSTDESVKQSANVLKSKVFYVPEYREMCLILLNNFDEKKMSITYLKDLVETTHVLIKLMEQFCGKSKHI
LTKKVIKKPNKNKKKAKQNKVLEPTEEELHGQWNEMSSELSAIIQGEADLPETVPFDAVSDQTEDEQKEISMRKI
NLLLRKKEMAEAVSLMRSSREVWPEGDIFGAGGLSPEDEFMALREVFMANLNPIEQQPVQEEELNGFDDEDEEDE
EDQYEERETELKIDEFIRRFANIKVMMAYSKLLDQYEKNSKFTNHCIMKMCHRIGFDAKLPAMFYQASIFRVFLK
AMEDPKGDFNESVKEIARFGKYIVRQFFTTAETNSKVFVELLFWKNNKEATEIECGYDDSSTKSSVKAASWLEEE
EDELQRLHGEFRFMEERPEGKDVLDLIEENMINKRRTRRQIVAKLKDMGLINNAQELKRKSVKVKGPKTWLDEEE
EELKMLVEEYKDHNNPMGMIMDMKVIKRPKQRVIDKILEMGLVNDSSELKKKRVLKVKKPKKSKKDPGLGDAFEM
ANQGDSDSDEDIVDDWSDSDEEKSKKPAKRIGADGKIKAKQSKKKDDSDTSDDESDDNQSKPTSSKAPYRPPVAT
PALISSALKTVKDAGMEEPIGWLSSLLKDTADDREEDGEFEPVPILALSESCTDAMEDSNFQNLLKLIGIHPPVT
HQEMFWRIPSRLTVEGLRKRAQYLSQGLEGIQIADEDADTPLEPTQERLSTQKTKKKKKTPKKTTKNKRKNKVVS
DEENIDPDDKENISVSRNDSSDDDDDDRPLSRLSSAPSASQNSINNDSKGETKKKGKKLSKKDFSTKEESDGDES
KLTKKRRFVINDDSDDETDMMDSHDLKLHLDTEPTQETNKSRIIDSDDENDQPSTKQTSKRLIESDSEDDVPLTS
RKKAKILDDSDED
```

Figure 2.5: Peptide sequence of *Euphausia superba* TIMEOUT. Black = TIMELESS domain. Purple = TIMELESS C.

2.3.2.5 Comparative analysis

Figure 2.6 shows schematic representations of the core circadian *Euphausia* proteins in comparison with *D. melanogaster* and *M. musculus* orthologs. Table 2.6 shows the identity/similarity scores generated from alignment of whole proteins and selected

domains. Going by alignment scores, EsBMAL1 showed a greater overall match to mBMAL1 than to dCYCLE, and three of the four comparable domains also favoured the mammalian ortholog. EsCLOCK was closer to dCLOCK overall, but comparable domains were equally split between dCLOCK and mCLOCK. EsPERIOD showed a greater overall similarity to dPERIOD, and two of the three comparable domains also received better scores against dPERIOD. As mTIMELESS is a paralog of dTIMEOUT, a comparison with EsTIMELESS was not considered appropriate.

2.3.3   Tissue expression

Esbmal1, Esclock, *Esperiod* and *Estimeless* were all confirmed by PCR to be expressed in antennal, eye, head, thorax and abdominal tissue (Figure 2.7). This is in line with previous work on *Escry1* and *Escry2* which showed both cryptochromes to be expressed in those tissues (Özkaya *et al*, unpublished).
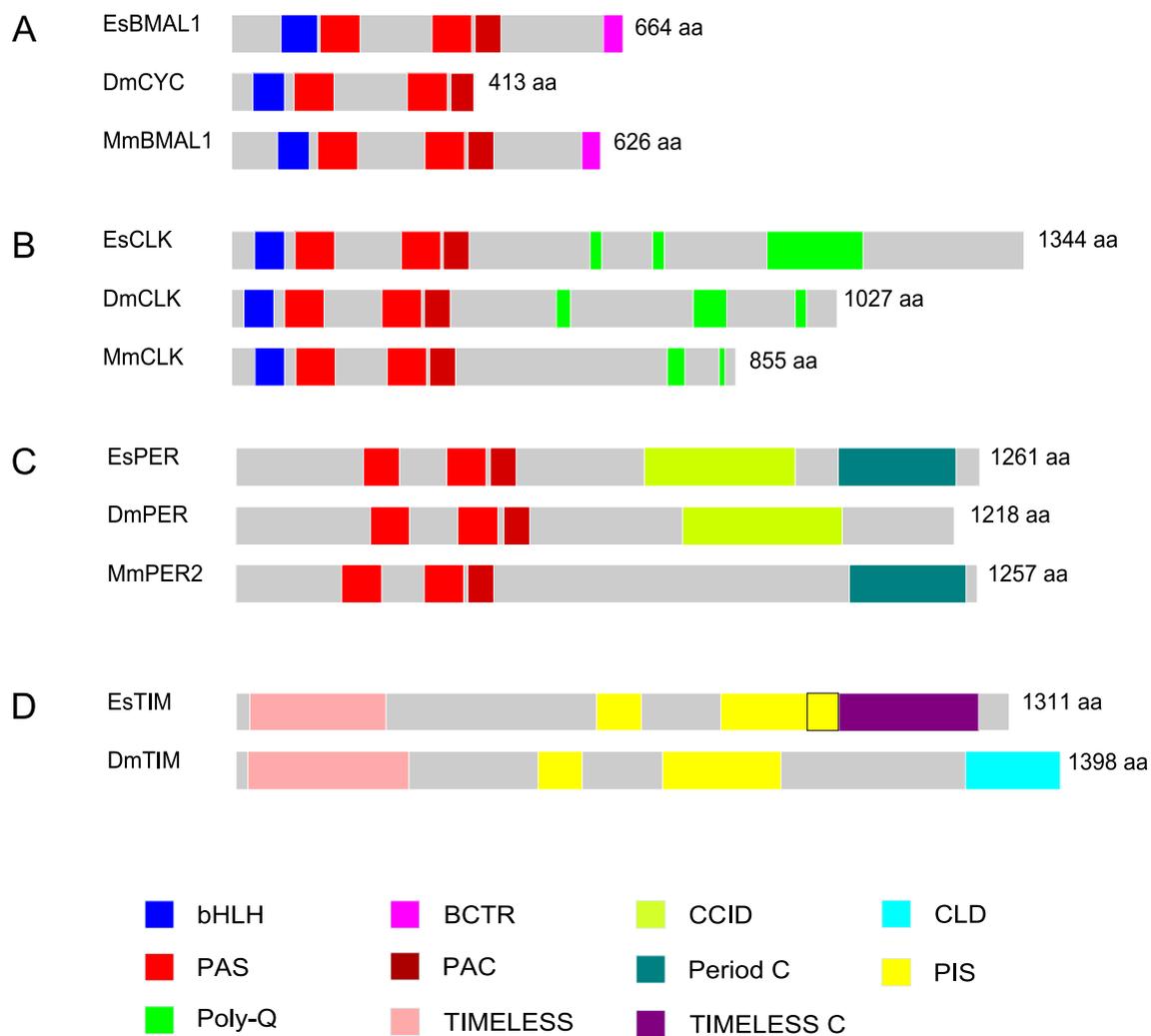
Figure 2.6: Schematic comparisons of canonical circadian proteins A) BMAL1/CYCLE, B) CLOCK, C) PERIOD and D) TIMELESS in *Euphausia superba*, *Drosophila melanogaster* and *Mus musculus*. TIMELESS in the latter was omitted due to its paralogous nature. Note: black outline on EsTIM indicates overlap of PIS and TIMELESS C.

Table 2.6: Identity and similarity of *Euphausia* circadian peptides and selected domains to equivalent sequences in *Drosophila melanogaster* and *Mus musculus*. Bolded entries show which species' sequence received the highest alignment score.

	versus *Drosophila melanogaster*			versus *Mus musculus*		
EsBMAL1	Identity	Similarity	Score	Identity	Similarity	Score
Whole protein	34	44.4	1146	43.5	57.1	**1388.5**
bHLH	74.1	85.2	187.5	72.2	85.2	**192.5**
PAS-A	69.1	85.3	244	73.5	89.7	**261**
PAS-B	50	79.4	202	55.5	74.6	**206**
PAC	61.4	75	**157**	47.7	77.3	134
BCTR	**-**	-	-	77.5	82.5	160
EsCLK						
Whole protein	26.3	34.9	**1439**	24.8	34.5	1229.5
bHLH	66.7	86.3	**180**	60.8	74.5	152
PAS-A	46.3	58.2	154	43.3	64.2	**155**
PAS-B	74.6	85.1	**291**	76.1	85.1	287
PAC	75	95.5	212	88.6	97.7	**237**
EsPER						
Whole protein	22.6	34.3	**898.5**	20.9	31.6	540.5
PAS-A	44	56	**154**	20.5	41.1	45
PAS-B	42	60.9	139.5	44.8	56.7	**144**
PAC	63.6	75	**162**	45.5	63.6	117
CCID	22.1	33.2	125	-	-	-
PERIOD C	-	-	-	18.4	29.6	82
EsTIM						
Whole protein	23.6	37.8	1466.5	-	-	-
TIMELESS	33.2	54	433.5	-	-	-
PIS-1	39.2	59.5	134	-	-	-
PIS-2	39.3	56.9	377.5	-	-	-
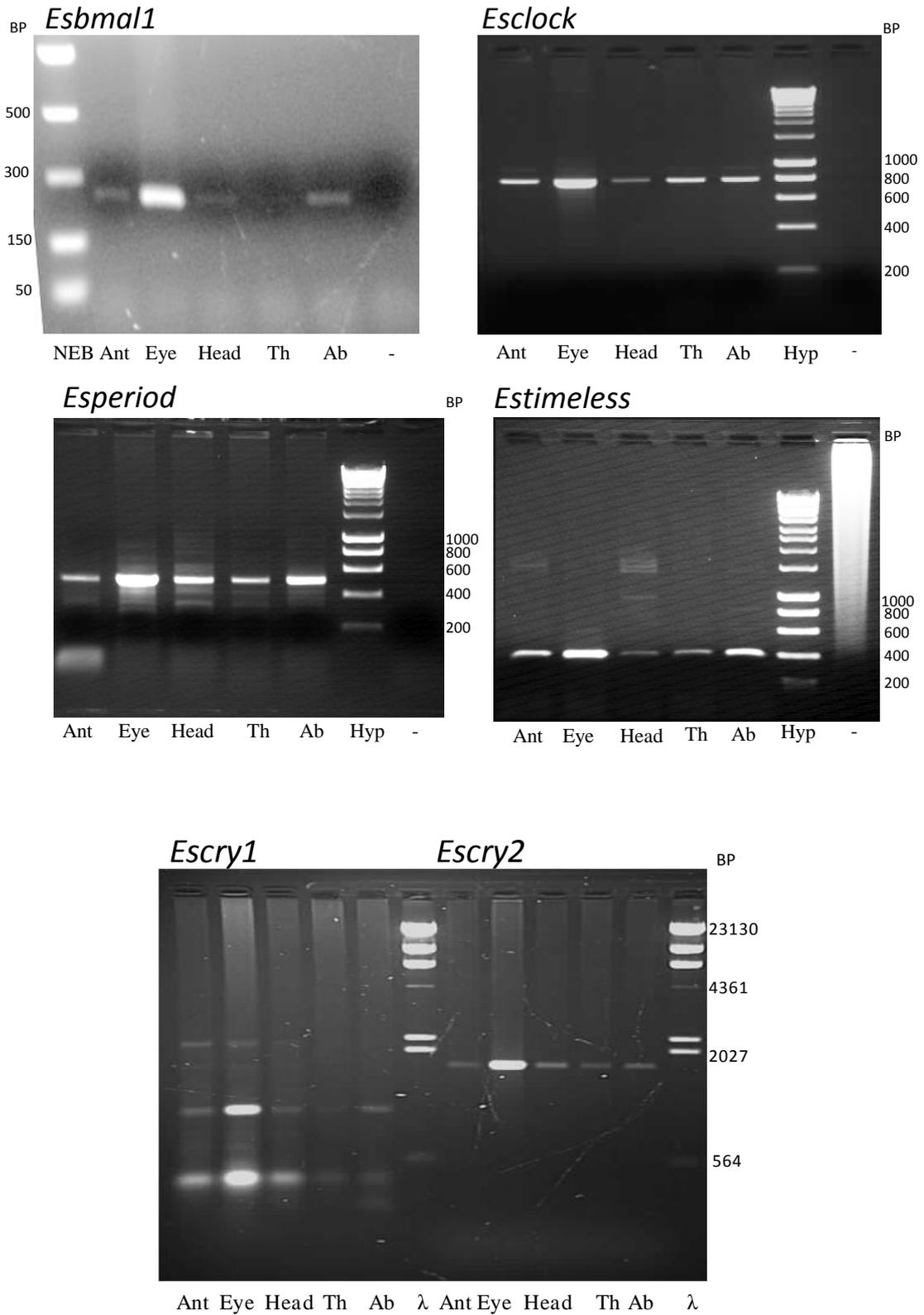
Figure 2.7: Tissue expression of the core clock genes of *Euphausia superba*. Ant = antennae. Th = thorax. Ab = abdomen. "-" = water control. *Esbmal1*, *Escry1* and *Escry2* results used with permission (Özkaya *et al*, unpublished). Ladders: BP = base pairs. NEB = PCR Marker NEB cat. N3234S. Hyp = Hyperladder I. λ – Lambda Hind III Digest. Note: Control reaction for *Estimeless* appears to have been contaminated with genomic DNA of unknown origin. Given that the expected *Estimeless* band is absent from this lane it was considered an acceptable control. See Appendix IV for further evidence.

2.3.4 Phylogenetic trees

Optimal and consensus neighbour-joining trees were generated for each circadian peptide. Bootstrap values represent the percentage of replicate trees in which the taxa clustered as shown in that branch, with only values > 50% shown. In the consensus tree, branches that were generated in less than 50% of bootstrap replicates are collapsed. While TIMEOUT is not considered a circadian component, it was included in the TIMELESS tree as a method of confirming the relationship of *Estimeless* and *Estimeout* to the two paralogous genes.

2.3.4.1 EsBMAL1

The optimal and consensus trees generated from BMAL1/CYCLE sequences are shown in Figure 2.8. With the exception of *Daphnia pulex* the crustacean entries were found at or near the base of the tree, the copepod orthologs in particular grouping near the annelid *Platynereis dumerilii* at the base of the vertebrate branch while the others rooted the tree. Exclusion of *Daphnia pulex* from the underlying alignment did not otherwise change the relationships depicted in the tree.

2.3.4.2 EsCLOCK

The optimal and consensus trees generated from CLOCK sequences are shown in Figure 2.9. The crustacean sequences grouped near *Platynereis dumerilii* at the base of the vertebrate branch, again with *Daphnia pulex* the exception. In the consensus tree *Daphnia* sits at the base of the crustacean group.

2.3.4.3 EsPERIOD

The optimal and consensus trees generated from PERIOD sequences are shown in Figure 2.10. The crustacean sequences grouped near the base of the invertebrate branch, above the sea snail *Bulla gouldiana*.

2.3.4.4 EsTIMELESS

The optimal and consensus trees generated from TIMELESS and TIMEOUT sequences are shown in Figure 2.11. The crustacean sequences group at the base of the invertebrate branches for each of the paralogs, and the trees confirm EsTIMELESS as a true TIMELESS rather than a TIMEOUT.
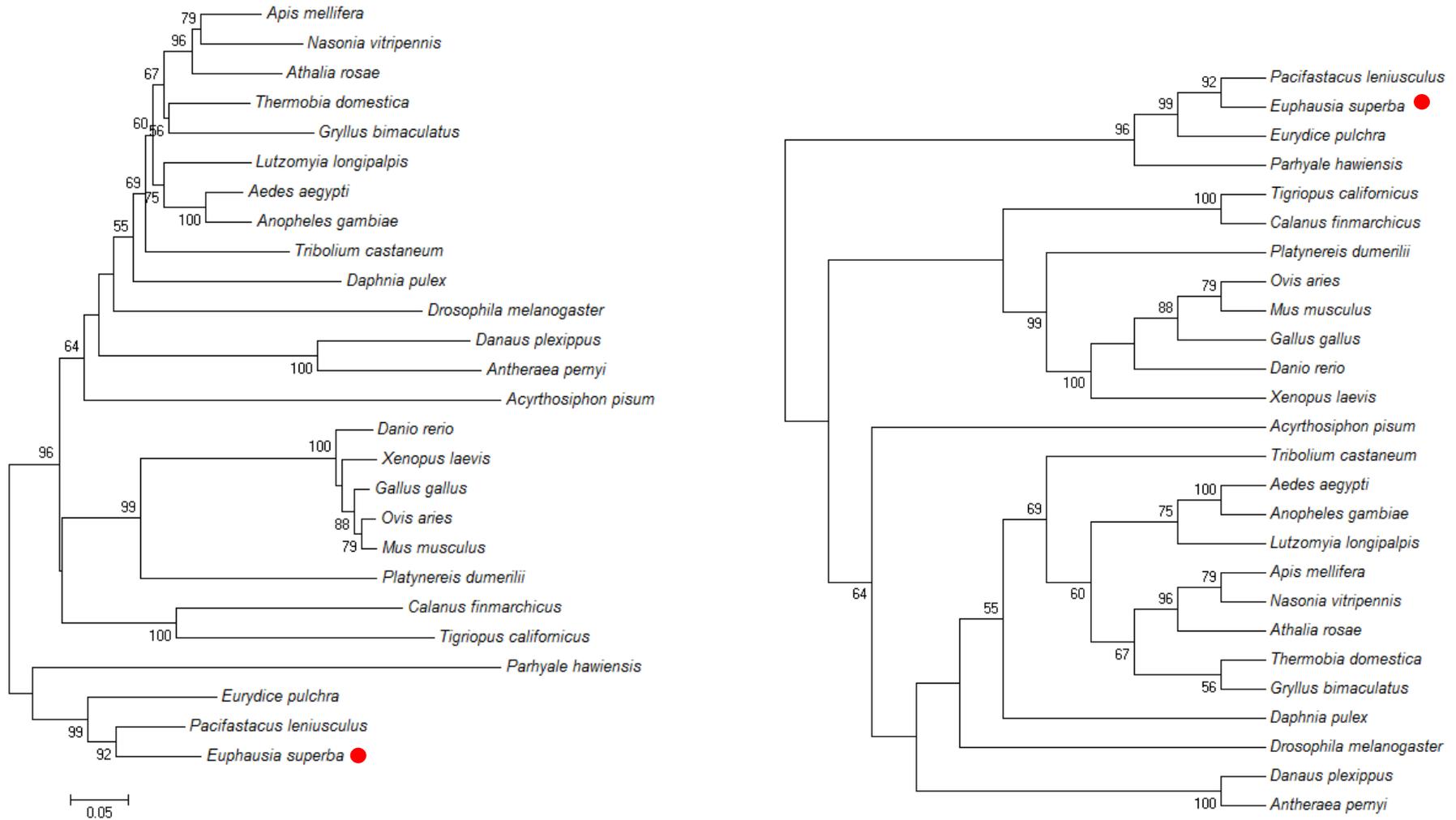
Figure 2.8: Optimal (left) and consensus (right) phylogenetic trees generated using BMAL1/CYCLE peptide sequences. Only bootstrap values over 50 are shown.
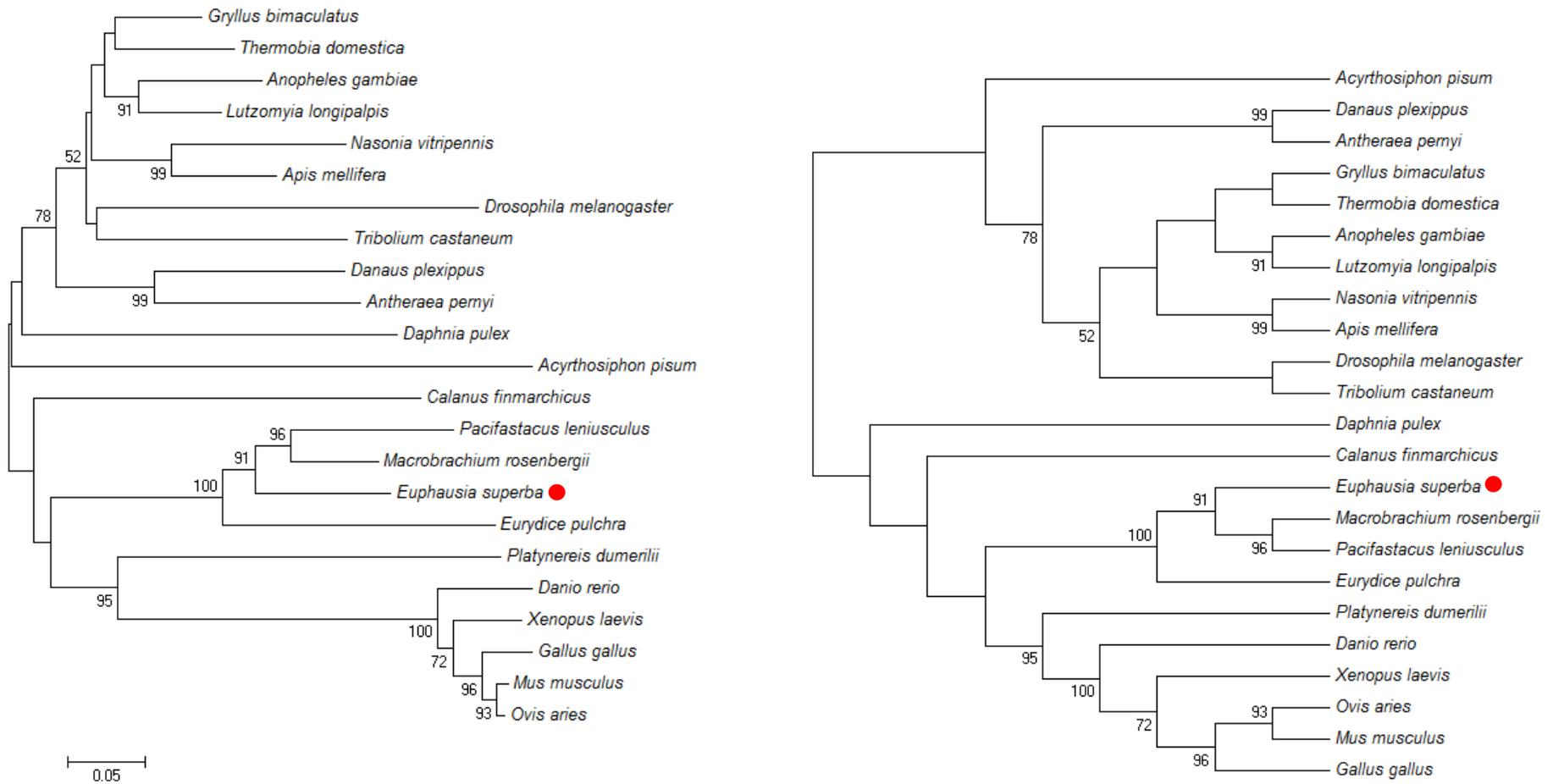
Figure 2.9: Optimal (left) and consensus (right) phylogenetic trees generated using CLOCK peptide sequences. Only bootstrap values over 50 are shown
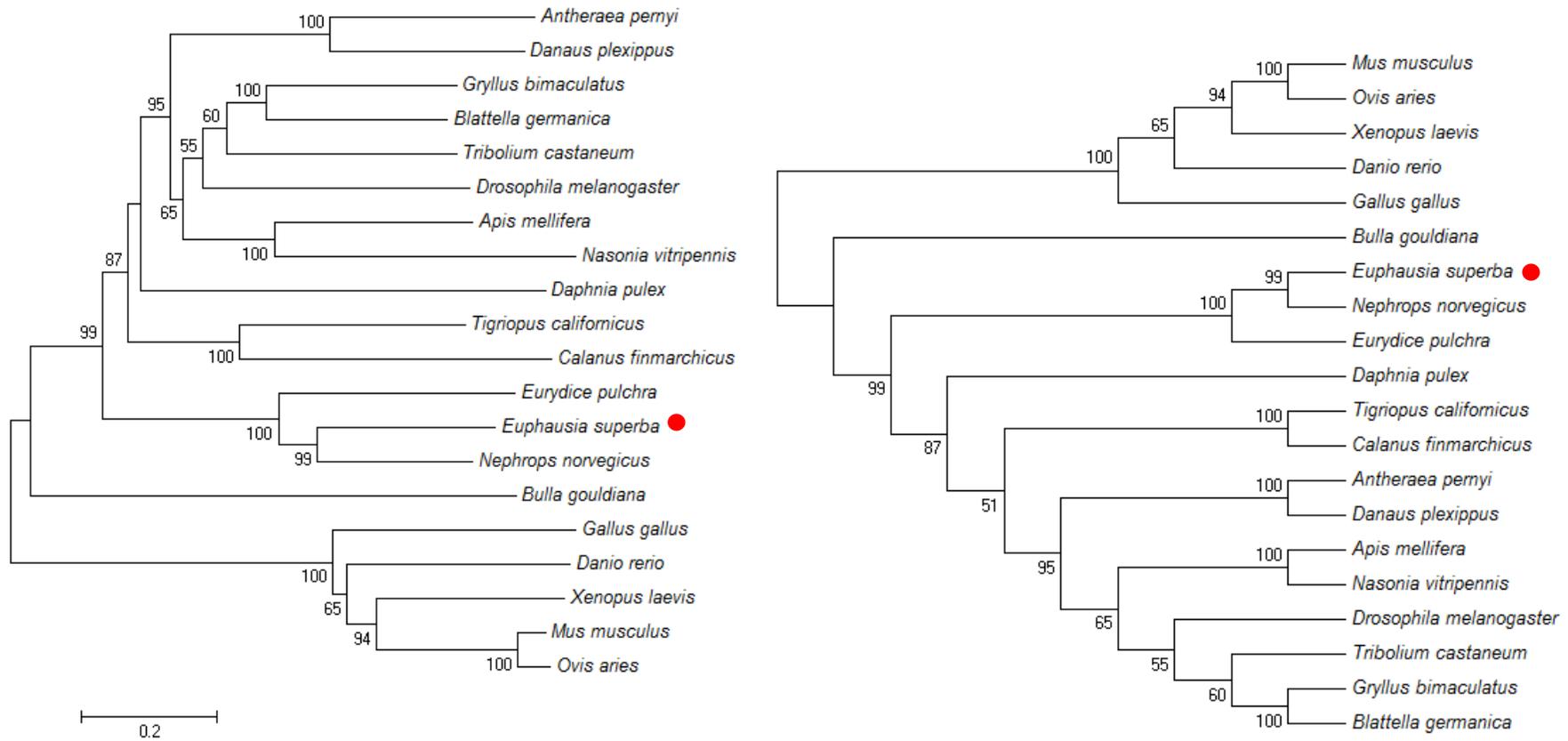
Figure 2.10: Optimal (left) and consensus (right) phylogenetic trees generated using PERIOD peptide sequences. Only bootstrap values over 50 are shown
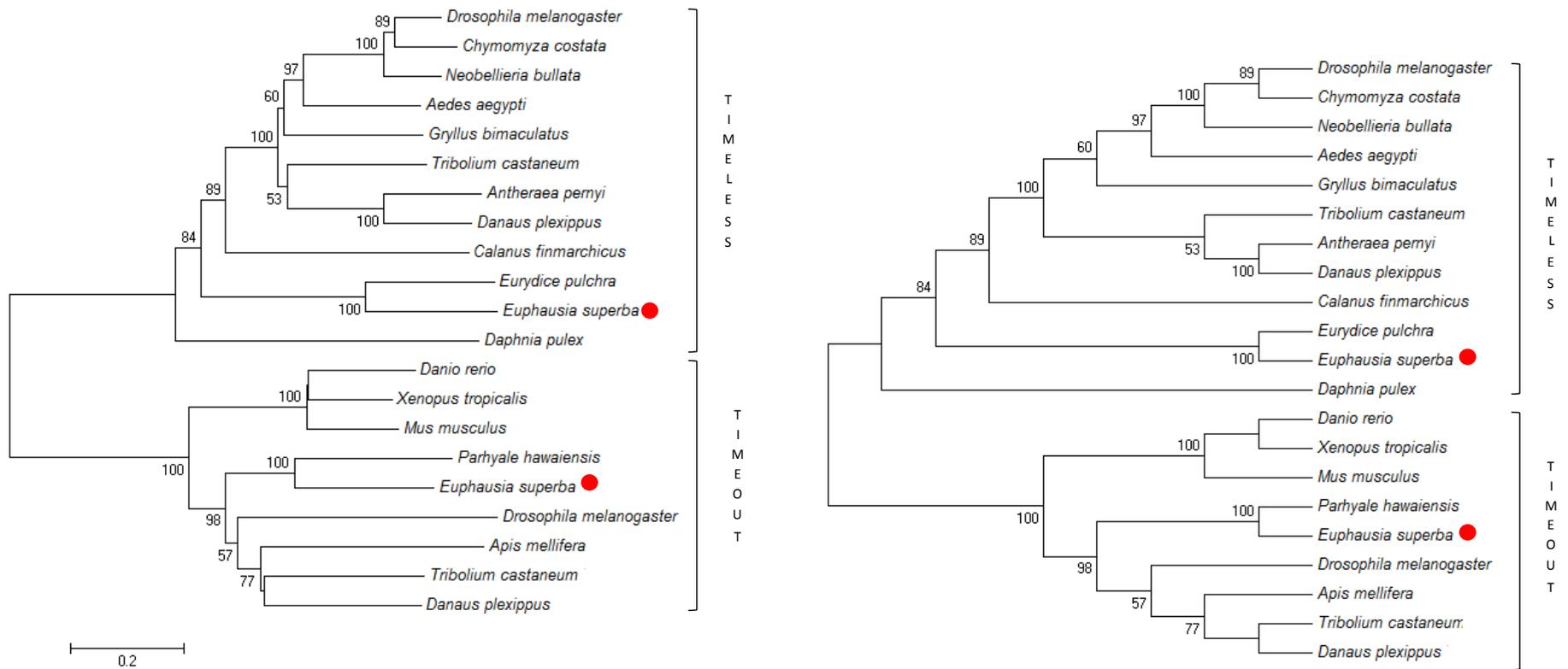
Figure 2.11: Optimal (left) and consensus (right) phylogenetic trees generated using TIMELESS/TIMEOUT peptide sequences. Vertebrate TIMELESS is termed TIMEOUT for the purposes of the tree. Only bootstrap values over 50 are shown.

2.3.5 Transcriptional activity

The relative luciferase activity of transfected S2 cells is shown in Figure 2.12. A one-way ANOVA returned a highly significant difference between transfection groups ($F_{7,16}$ = 556.4, $p < 2e10^{-16}$), with a Tukey post-hoc test used to confirm in which groups this difference lay. Cotransfection and expression of *Esclock* and *Esbmal1* into S2 cells was necessary to see luciferase activity beyond background (p <0.001 against empty pAc5.1 plasmid control), with neither alone proving sufficient to significantly drive expression of the luciferase reporter; relative expression in both cases was no different to that of the empty plasmid control (*Esbmal1* only, p = 1, *Esclk* only, p = 1).

Cotransfection of *Esper* in cells transfected with *Esclk* and *Esbmal1* significantly reduced luciferase activity to approximately 60% (p < 0.001 against *Esclk:Esbmal1* transfected cells). The same was true for *Estim* (p < 0.001). Cotransfection of both *Esper* and *Estim*, meanwhile, proved a much more potent repressor of luciferase activity than either alone (p < 0.001). Finally, cotransfection with *Escry2* again showed a repressive effect (p < 0.001 against *Esclk:Esbmal1* transfected cells) that was as potent as *Esper* and *Estim* together (p = 0.99).

EsBMAL1	+	+	-	-	+	+	+	+
EsCLK	+	-	+	-	+	+	+	+
EsPER	-	-	-	-	+	-	+	-
EsTIM	-	-	-	-	-	+	+	-
EsCRY2	-	-	-	-	-	-	-	+
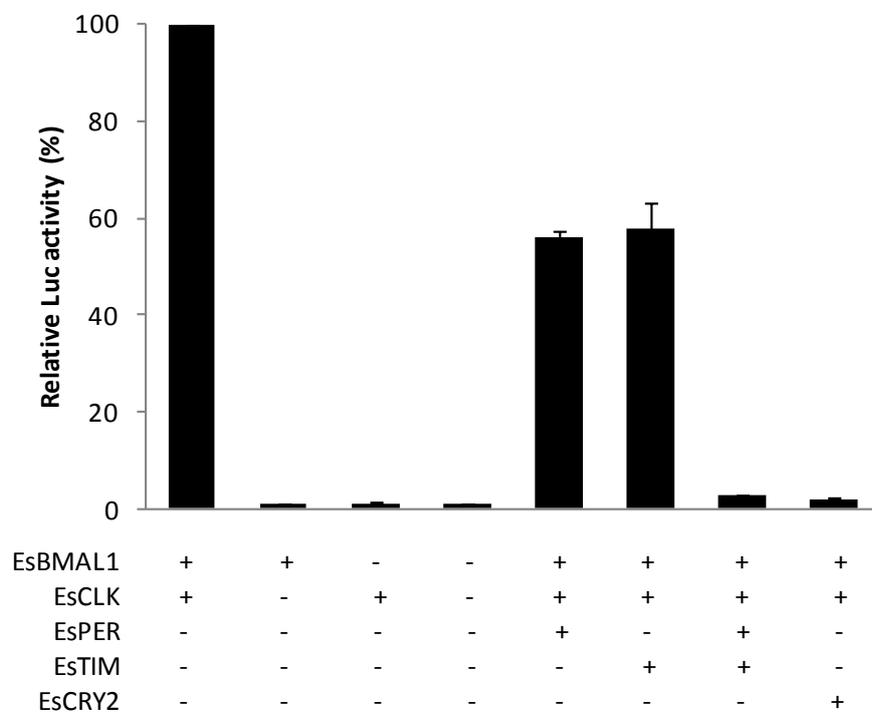
Figure 2.12: Mean relative E-box driven luciferase activity (+ SD), normalised to *Renilla*. Three independent transfections were performed for each assay.

2.4   Discussion

2.4.1   Summary of results

2.4.1.1 *Esbmal1*

A 2,017 bp coding sequence encoding a 664 amino acid peptide was successfully cloned and confirmed by alignment and domain analysis to be the *Euphausia superba* ortholog of *D. melanogaster cycle*. Given the presence of a BCTR the gene was named for the *Mus musculus* ortholog *bmal1*. Pairwise alignment of whole peptides and selected domains with *D.melanogaster* and *M. musculus* orthologs suggested a greater similarity to the latter. Phylogenetic analysis shows EsBMAL1 near the root of a neighbour-joining tree, grouped with other crustaceans, albeit not all that were included.

Esbmal1 appears to be expressed in the antennae, eye, head, thorax and abdomen of *Euphausia.* The protein product EsBMAL1 is necessary, but not sufficient to drive the expression of an E-box-luciferase reporter in embryonic S2 cells, requiring the presence of EsCLOCK.

2.4.1.2 *Esclock*

A 4,032 bp coding sequence encoding a 1,344 aa peptide was cloned and confirmed to be the *Euphausia superba* ortholog of *D. melanogaster Clock*. Domain analysis revealed a large poly-glutamine region in the C-terminal region. Pairwise alignment of whole peptides and selected domains with *D.melanogaster* and *M. musculus* orthologs suggested an overall greater similarity to the former, although domain scores were equally divided between the two. Phylogenetic analysis of the peptide groups EsCLOCK with other crustaceans at the base of the vertebrate branch.

Esclock appears to be expressed in the antennae, eye, head, thorax and abdomen. EsCLOCK is necessary, but not sufficient, to drive the expression of an E-box-luciferase reporter in embryonic S2 cells, requiring the presence of EsBMAL1.

2.4.1.3 *Esperiod*

A 3,783 bp coding sequence encoding a 1,261 aa peptide was cloned and confirmed to be the *Euphausia superba* ortholog of *D. melanogaster period*. Phylogenetic analysis of

the peptide EsPERIOD places it with other crustaceans at the base of the invertebrate branch. Pairwise alignment of whole peptides and selected domains with *D.melanogaster* and *M. musculus* orthologs suggested a greater similarity to the former.

Esperiod appears to be expressed in the antennae, eye, head, thorax and abdomen. EsPERIOD is capable of significantly repressing EsBMAL1:EsCLOCK-mediated expression of an E-box-luciferase reporter to around 60%, but when cotransfected with EsTIMELESS is a much more potent repressor.

2.4.1.4 *Estimeless*

A 3,933 bp coding sequence encoding a 1,311 aa peptide was cloned and confirmed to be the *Euphausia superba* ortholog of *D. melanogaster timeless*. Phylogenetic analysis confirmed it to encode a true TIMELESS rather than the paralogous TIMEOUT (also cloned during this work), and placed it at the base of the invertebrate branch with other crustaceans.

Estimeless appears to be expressed in the antennae, eye, head, thorax and abdomen. EsTIMELESS is capable of significantly repressing EsBMAL1:EsCLOCK-mediated expression of an E-box-luciferase reporter to around 60%, but when cotransfected with EsPERIOD is a much more potent repressor.

2.4.1.5 The cryptochromes

In addition to the circadian components described above for the first time, *Euphausia superba* possesses two *cryptochrome* genes; not only the mammalian-like *Escry2* (Mazzotta *et al.*, 2010) but also a *Drosophila*-like *Escry1* that has been confirmed to be light-sensitive (Figure 2.13; Özkaya *et al*, unpublished). In this work, EsCRY2 has been shown to be a potent repressor of EsBMAL1:EsCLOCK mediated luciferase reporter expression.
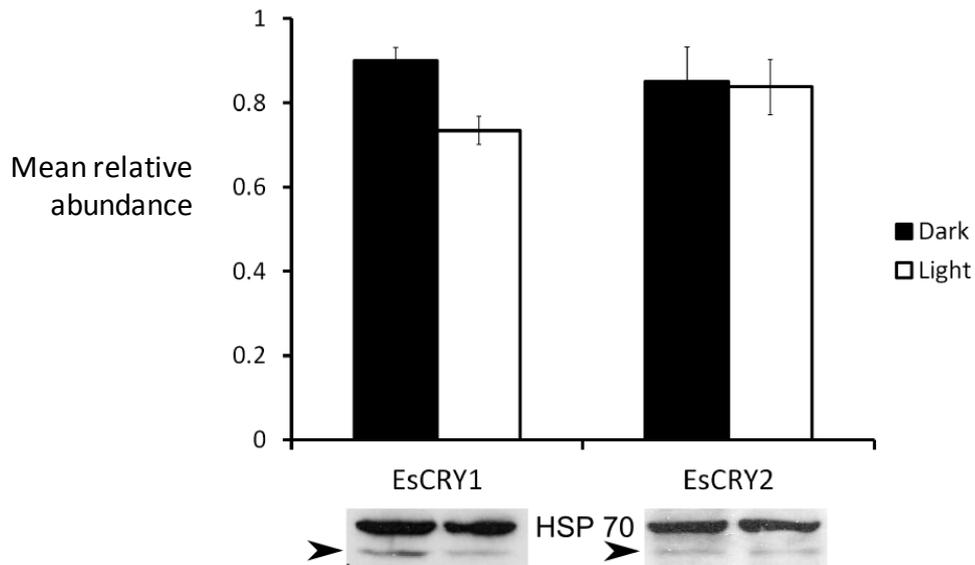
Figure 2.13: Degradation of *Euphausia* CRY1 protein on exposure to light. Transgenic flies possessing *Escry1* or *Escry2* were used to assess photosensitivity through Western blot of protein products after maintenance in dark or light conditions. Mean relative abundance shown, normalised to HSP70 abundance. Aligned below their respective chart columns are the Western blot results - arrows indicate EsCRY1 and 2 abundance. Error bars +/- SE. Used with permission from Özkaya *et al* (unpublished).

2.4.2 Peptide features and comparisons

2.4.2.1 The core heterodimer

Both components of the heterodimer possess the typical bHLH and PAS domains that enable DNA-binding and dimerisation respectively. In line with the majority of species studied, EsBMAL1 possesses a highly-conserved BCTR. Given the presence of this region in both vertebrates and invertebrates alike it has been suggested previously that this may be the ancestral form, and that dCYC is a derived type that has secondarily lost the C-terminus (Chang *et al.*, 2003). The results shown here uncontroversially support this assertion, with multiple representatives from Crustacea now also shown to possess a BCTR. As Insecta evolved from within Pancrustacea (Regier *et al.*, 2010) and a BCTR has been found in many insects, the support for this hypothesis can be considered strengthened.

EsCLK is the longest *Clock* protein identified so far, containing the longest poly-Q region. Another malacostracan crustacean, the giant river prawn *Macrobrachium rosenbergii*, also has an exceptionally long poly-Q region (Yang *et al.*, 2006), but while

an interesting observation this does not appear to be typical of crustaceans (Table 2.7). It is interesting to note however that the peptides for *Daphnia pulex* and *Calanus finmarchicus*, neither of which show any sign of a poly-Q, come from genome/transcriptome mining and have not been further confirmed by PCR or RACE. It is therefore hard to draw an informed conclusion on them given the often partial nature of transcriptome sequences and the reliance on assembly quality for predicted peptides from genomes. A cautionary tale on this very subject is seen with reference to Chang *et al.* (2003) who speculated that poly-Q regions may have risen independently in mammals and *Drosophila* due to the lack of such regions in the orthologs of *Antheraea pernyi* and the *Anopheles gambiae*. While the former sequence was obtained through degenerate PCR and RACE extension and so can be considered complete, the mosquito CLK (accession EAA11642) was retrieved from the genome and at the time did indeed lack a poly-Q. The sequence was updated in 2011, however, and the current version (EAA11642.4) contains multiple such regions including a 58 residue stretch in which 35 are glutamine. Given the presence of extensive poly-Q regions in Crustacea and Annelida (Table 2.7) it seems more likely that the absence of these in certain species represents a departure from an ancient type in much the same way as the loss of the BCTR in dCYC. We will return to this evolutionary aspect in Chapter Six.

Is there any function to the CLOCK poly-Q in species other than *Drosophila*? Yang *et al.* (2006) speculate that it may act as a transcriptional activator, citing evidence that poly-Q expansion increases transcription activity of the GAL4 transcription factor *in vitro*, and that such regions are common to transcription factors (Gerber *et al.*, 1994). If so it is not sufficient to do so in *Euphausia*, EsCLK requiring the presence of EsBMAL1, and given the presence of a highly conserved BCTR in the latter, perhaps it is likely that the poly-Q functions in modulation of transcription allowing fine control, or supporting the activity as is seen with mice (Takahata *et al.*, 2000).

Table 2.7: Length of largest glutamine-rich regions and count of glutamine residues contained within for various species. Orange = Crustacea. Blue = Insecta. Yellow = Chordata. Purple = Annelida. Adapted from Yang *et al*, (2006).

Organism	Q-rich region/Qs (amino acids)
Euphausia superba	163/116
Macrobrachium rosenbergii	140/135
Drosophila melanogaster	57/47
Danio rerio	51/43
Platynereis dumerilii	40/28
Eurydice pulchra	32/28
Gallus gallus	20/18
Mus musculus	19/17
Xenopus laevis	14/8
Antheraea pernyi	None
Calanus finmarchicus	None
Daphnia pulex	None

Another possibility presents itself when considering the latitudinal clines of poly-Q regions seen in vertebrate CLOCKs as identified in the blue tit *Cyanistes caeruleus* (Johnsen *et al.*, 2007) and the Chinook salmon *Oncorhynchus tshawytscha* (O'Malley and Banks, 2008). In both species there is a correlation between poly-Q length and latitude, with longer alleles found more frequently at higher latitudes, suggesting that the variation may represent local adaption to photoperiod or temperature. In the blue tit, the poly-Q variation in a single population has also been linked with seasonal phenotypes and local adaptation, with short poly-Q birds breeding earlier, incubating eggs for a shorter length of time and having greater reproductive success (Liedvogel *et al.*, 2009). It is tempting to suggest that as a species found at the highest of latitudes and subject to extreme seasonal variations *Euphausia*'s poly-Q represents the apotheosis of this pattern, but the idea is rather undermined by the extensive repeats found in the tropical *Macrobrachium rosenbergii*.

A similar phenomenon is seen in the circadian clock of *Neurospora crassa,* in which the heterodimer that drives transcription is the White Collar Complex (WCC), comprised of WC-1 and WC-2 which bind via PAS domains (Heintzen and Liu, 2007). At the N-terminus of WC-1 is a poly-Q, loss of which renders the fungus arrhythmic and the length of which varies between strains. Poly-Q length and circadian period are correlated, longer repeats resulting in longer periods, again linked to a latitudinal

pattern (Michael *et al.*, 2007). This is not to suggest direct homology between the fungus and krill clocks of course, but to emphasise again the type of circadian phenomena that CAG repeats are associated with. Overall, the EsCLK poly-Q is a notable feature worthy of further investigation.

2.4.2.2 The domains of EsPERIOD and EsTIMELESS

To continue the theme of trinucleotide repeats, a Thr-Gly length polymorphism in *Drosophila* PERIOD has been identified to underlie circadian temperature compensation, with the alleles distributed in a latitudinal cline (Sawyer *et al.*, 1997). No such feature stood out in EsPERIOD however, despite the presence of three single Thr-Gly pairs in the peptide, and it remains a distinctly Drosophilidae phenomenon (Nielsen *et al.*, 1994).

The CCID of dPERIOD is implicated in repression of the dCLK:dCYC heterodimer in *Drosophila* (Chang and Reppert, 2003) but the low degree of identity/similarity compared to other conserved domains (Table 2.6) means the putative *Euphausia* CCID should not be considered functional or confirmed present without experimental evidence. Similar caution should be taken with the PER-interaction sites of EsTIMELESS, which were again identified only through alignment, albeit with better alignment scores. Both the CCID and PIS are in theory supported by the results of the S2 cell assay (see 2.4.3.1), but again direct evidence pinpointing sites of interaction should be sought.

The failure to identify a CLD in EsTIMELESS is unsurprising, and should not be considered evidence that it is not present. The domain is poorly conserved even amongst Drosophilids with only 56% identity between *Drosophila melanogaster* and *Drosophila virilis* (Ousley *et al.*, 1998) and identification will again rely on functional analysis rather than alignment.

One other point of interest is the incomplete nature of the TIMELESS and TIMELESS C domains in EsTIMELESS as specified by SMART analysis. These domains are representative of the paralogous proteins but without clearly defined functions akin to that of a PAS or bHLH, and the apparent deterioration of them in the 'true' TIMELESS of *Euphausia* stands in contrast with the finding that of the two, *timeless* appears to be

under stronger purifying selection (Gu *et al.*, 2014), although this could simply be an artefact of the sequences chosen to define the domains.

2.4.2.3 Evolutionary relationships

The neighbour-joining trees generated in this work are by no means offered as the final word on the phylogeny of the canonical clock genes – if nothing else, the failure of certain crustacean sequences to group together tells us not to draw any firm evolutionary conclusions from them – but some inferences can be taken in conjunction with the comparative domain analysis.

In the case of BMAL1/CYC, regardless of the deletion parameters and sequences used the placement of the crustacean sequences was reliably either at the base of the entire tree or grouping with the vertebrate branch, though the nodes were never particularly stable. The comparison of the whole protein and bHLH, PAS, PAC and BCTR domains also favoured *Mus* BMAL1 over the *Drosophila* ortholog, suggesting that the Insecta branch may be more derived. Similarly with EsCLK, the crustacean peptides showed greater similarity to their vertebrate counterparts according to the generated trees, although protein and domain alignment slightly favoured dCLK in this case. EsPER and EsTIM show less remarkable patterns, stably grouping with the invertebrates in both cases. Overall, a picture forms of a core heterodimer in particular that retains ancestral characteristics.

A moment should also be taken to cover two other particular features. First is the occasionally eccentric placement of *Daphnia pulex*. As an animal with the ability to reproduce parthenogenetically (Innes *et al.*, 1986), it is perhaps a genetic oddity that may on occasion throw out an otherwise sensible alignment. It should be noted that the placement is restricted to the optimal trees, and in the consensus trees it is more reasonably sited. Secondly is the well-supported placement of the CLK (95% of bootstrap replicates) and BMAL1 (99%) of the annelid *Platynereis dumerilii* at the base of the vertebrate orthologs, near the crustaceans. Considering the modern consensus phylogeny of Bilateria (Edgecombe *et al.*, 2011) which places Spiralia/Lophotrochozoa (containing Annelida) as a sister group to Ecdysozoa (containing Arthropoda) that diverged after the split from Deuterostomia (containing Vetebrata), this might be

considered further evidence of the particularly derived nature of the Insecta orthologs. As more orthologs are gathered from un/underrepresented taxons, the relationships will hopefully clarify further.

2.4.3 The molecular clockwork of *Euphausia superba*

2.4.3.1 Transcription, repression and tissue expression

It has been shown here that *Euphausia*'s circadian components EsCLK and EsBMAL1 are both required to drive expression of E-box sequences, as has been found in every other animal studied in similar assays. As *Drosophila cycle* is expressed in S2 cells (Darlington *et al.*, 1998), a further implication is that the EsCLK is unable to dimerise with CYC and requires the ortholog of its own species – or at least one evolutionarily closer – to pair with and/or subsequently function.

As EsBMAL1 has a BCTR, and bearing in mind the evidence from assays using *Antheraea pernyi, Danaus plexippus, Eurydice pulchra* and *Mus musculus* orthologs, it was expected that EsCRY2 would be capable of repressing transcription, though the completeness of the repressive effect at a transfection amount ratio of 1:1:1 (100 ng of each component) was surprising when compared to the results in other species. In *Eurydice pulchra*, for example, 50 ng of *Epcry2* is required to reduce the effect of 1 ng each of *Epbmal1* and *Epclk* to ~10% (Zhang *et al.*, 2013), while in *Danaus plexippus* the ratios are the same as those used here (50 ng of each component) but *Dpcry2* reduces reporter activity to just under 20% (Zhu *et al.*, 2008).

More surprising was the effect of cotransfection with *Estimeless* and *Esperiod*. Both show the ability to moderately repress the activity of the heterodimer individually, a result in line with findings in *Drosophila* (Lee *et al.*, 1999) and the more closely related *Eurydice pulchra* (Zhang *et al.*, 2013). When both are cotransfected however the repressive effect is much more potent, to the extent that there is no significant difference in expression to that of a negative control. The surprise lies not in their ability to do so, but in doing so as effectively as EsCRY2. In *Antheraea pernyi* apPER repress transactivation activity, albeit in a dose response manner requiring high doses for potent repression. At a ratio of 2:1:1 against the core heterodimer (2 ng apPER, 1 ng apCLK, 1 ng apBMAL), expression is reduced to ~75%; at 10:1:1 just under 50%; at

50:1:1 less than 25% and at 250:1:1 less than 10%. This effect is augmented slightly at each dosage by low dose cotransfection with apTIM. In *Danaus plexippus,* meanwhile*,* DpPER and DpTIM show no repressive capability, and are believed instead to form cytoplasmic complexes with DpCRY2 in order that the latter can be shuttled into the nucleus to enact its repressive effect (Zhu *et al.*, 2008).

Prior to performing the assay in *Euphausia*, the assumption was that a similar result to that of *Danaus plexippus* would be obtained, the model making logical use of all available components and their domains (Figure 2.14). The enhanced repression seen with cotransfection however is reminiscent of the repressive effect of the dPER:dTIM dimer (Darlington *et al.*, 1998), although again to a greater extent than seen in that study, and the implication is that *Euphausia*'s clockwork can function like that of *Drosophila* as well as showing the more common cryptochrome-based feedback loop.

Tissue-specific PCR indicates that all the clock genes are co-expressed in the tissues tested. While not a quantitative analysis, PCR in the eye tissue consistently produces the strongest band which suggests, in line with other crustaceans, that the eye may be the site of one of the major oscillators in *Euphausia* - the results of *Estimeless* in particular suggest very low expression in other tissues. Having established evidence of gene expression in these tissues it will be illuminating to proceed to finer dissections to establish if expression is restricted to certain sites such as the brain, ovary or hepatopancreas.
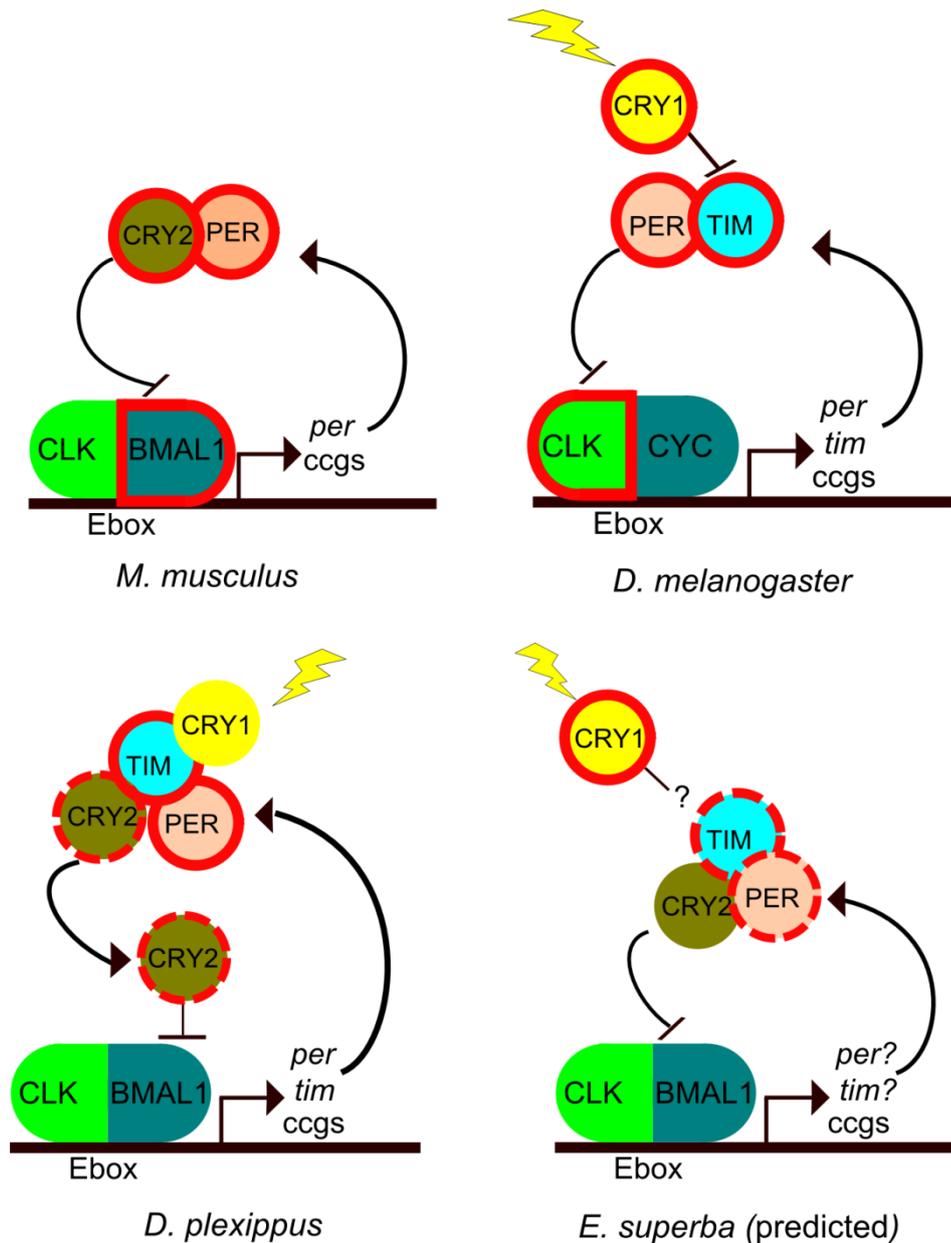
Figure 2.14: Models of the negative feedback loop in *M. musculus, D. melanogaster and D. plexippus*, and predicted model based on identified components of *Euphausia superba*. Red outline indicates a confirmed cycling transcript. Broken red outline indicates transcript cycling yet to be determined. Note: *D. plexippus* redrawn from Reppert (2007) – significant cycling of dpCRY2 was not found in head tissue RT-PCR, but was seen in DpN1 cells, hence broken outline.

2.4.3.2 An ancient clock or a crustacean specialisation?

The clock of *Danaus plexippus* has been suggested as a candidate representing the ancestral type (Zhu *et al.*, 2008), given the discovery of dpCLK, dpCYC, dpPER, dpTIM, dpCRY1 and dpCRY2 – that is, the core components of both the *Drosophila* and *Mus* clocks found in a single animal. The same can now be said for *Euphausia superba.* But

further to this, EsCLK shows what is increasingly looking like the ancestral, poly-Q form, whereas *Danaus plexippus* lacks such a region. Given that the same is true for *Antheraea pernyi*, this might be a Lepidopteran derivation.

Furthermore, phylogenetic analysis and domain comparisons of *Euphausia*'s circadian proteins to their *Drosophila* and mammalian counterparts are suggestive of a system that sits between the two, with the transcriptional activators favouring comparison with *Mus musculus* while the repressive elements EsPER and EsTIM have a more invertebrate character.

And finally, the clock of *Euphausia* is not simply in possession of the canonical fly and mouse components but also seemingly their respective mechanisms, at least *in vitro*. Complete repression of E-box mediated luciferase reporter expression by two separate paths has to our knowledge not been reported before in circadian studies, although it is important to note the context of the assay in using *Drosophila* S2 cells – *Danaus plexippus* PER shows the ability to repress dpCLK:dpCYC activity in a dosage dependent manner in the same cells, yet not in the more relevant environment of the butterfly's DpN1 embryonic cells (Zhu *et al.*, 2008).

Should it prove that *Euphausia* does indeed possess two repressive mechanisms, what are the implications? One idea is in line with the above – that *Euphausia* possesses an ancestral clock in which all options are viable, while subsequent insect and vertebrate evolution locked each into their own specialised pathways. This may be specific to *Euphausia*, as cotransfection of *Eurydice pulchra* PER and TIM together did not generate the same results (Zhang, personal communication).

Another suggestion is that one loop controls circadian phenomena while the other controls seasonal responses. In *Drosophila* diapause as a response to photoperiod is linked to TIM via the light-sensitive CRY (Sandrelli *et al.*, 2007), and *timeless* is similarly implicated in larval diapause in *Chymomzya costata* (Riihimaa and Kimura, 1988). Photoperiod has been shown to regulate gene expression in *Euphausia* (Seear *et al.*, 2009) and it undergoes a remarkable seasonal regression to a juvenile overwintering state (Kawaguchi *et al.*, 2007). Now shown to be in possession of a *Drosophila*-like TIM

and a light-sensitive CRY1, *Euphausia* may enact such responses through a dedicated pathway while the mammalian-like loop controls daily rhythms.

Finally, it could offer a molecular underpinning for the multiple circadian oscillators theory of crustacean chronobiology. *Carcinus maenas* can entrain to three separate circatidal zeitgebers at once (Warman and Naylor, 1995) and a central pacemaker is yet to be identified in any crustacean. Perhaps this ability to maintain multiple rhythms is enabled not only by independent, tissue-specific oscillators, but by oscillator-specific molecular clockwork. Two putative pathways to complete repression have been identified in *Euphausia*, and the components EsPER and EsTIM have also individually shown the ability to significantly repress expression, raising the possibility of fine-grained control when required. This particular hypothesis would have been bolstered by the identification of tissue-specific restriction of expression of certain components, but this did not prove to be the case.

2.4.3.3 Regarding transcript cycling

A notable absence in the results detailed in this chapter is any data on transcript cycling, a common feature of core clock genes. *Clock* is rhythmically expressed in *Drosophila* while *cycle* expression in constitutive, while the opposite is true in mice (Reppert and Weaver, 2000), and their actions drive the rhythmic expression of clock controlled genes including *period* and *timeless*. Quantitative real-time PCR (RT-PCR) assays on *Euphausia*'s *cryptochrome* genes and *Esclock* and *Esbmal1* have been conducted in our lab (Özkaya *et al*, unpublished; see Figure 2.15 for results and details), with only *Escry1* showing consistent evidence of transcript cycling in entrained samples taken from LD and DD conditions, and in wild caught samples.

The lack of cycling in *Escry2* stands in contrast to results obtained by others. Mazzotta *et al.* (2010) detected daily fluctuations of *Escry2* mRNA in wild caught samples, with the peak coming at 6 am, which is in agreement with our entrained LD results but not our own catch samples. Meanwhile Teschke *et al.* (2011) found rhythmic expression in krill kept in both LD and DD. The mRNA peak during the light phase came later than seen in the other study, at midday, while the DD results suggested a short endogenous period of approximately 18 hours, although the experiment was run twice and no

significant oscillations were detected in the second run. Our entrained DD results on the other hand show no evidence of cycling. All in all, the picture is distinctly unclear.

There are notable differences in methodology between our approach and that of Teschke *et al*: in the cited study individual experimental *Euphausia* were entrained for 24 hours before samples were taken, compared to 48 hours entrainment for our own assays. However Teschke *et al* also kept the wider stock of animals from which samples were taken in the same conditions for 3 – 4 weeks prior to sampling to allow acclimatisation to tank conditions, whereas ours were entrained direct from capture. Perhaps the most significant difference between the two approaches comes from the choice of control. Teschke *et al* chose to use *phosphoenol-pyruvatecarboxykinase (pep-ck)* as an internal reference gene against which to normalise transcript levels, a previous study (Seear *et al.*, 2009) having indicated stable mRNA levels across samples. Our study on the other hand made use of an *Aequorin* RNA spike-in control against which to normalise results.

Figure 2.15 (following page): RT-PCR expression of *Euphausia superba* genes *Esbmal1, Esclock, Escry1* and *Escry2*. Top – Entrained LD samples, ZT = zeitgeiber time, ZT0 = 4 am, black bar indicates dark phase. Middle – Entrained DD samples, CT = circadian time, grey bars indicate relative light phase during entrainment. Bottom – wild catch samples, actual time shown. Error bars = standard error. N = 3 per timepoint, each replicate generated from cDNA library of 3 krill heads. Expression values normalised to *Aequorin* spike-in control. Reproduced with permission from Özkaya *et al* (unpublished).

Esbmal1

Esclock

Escry1

Escry2
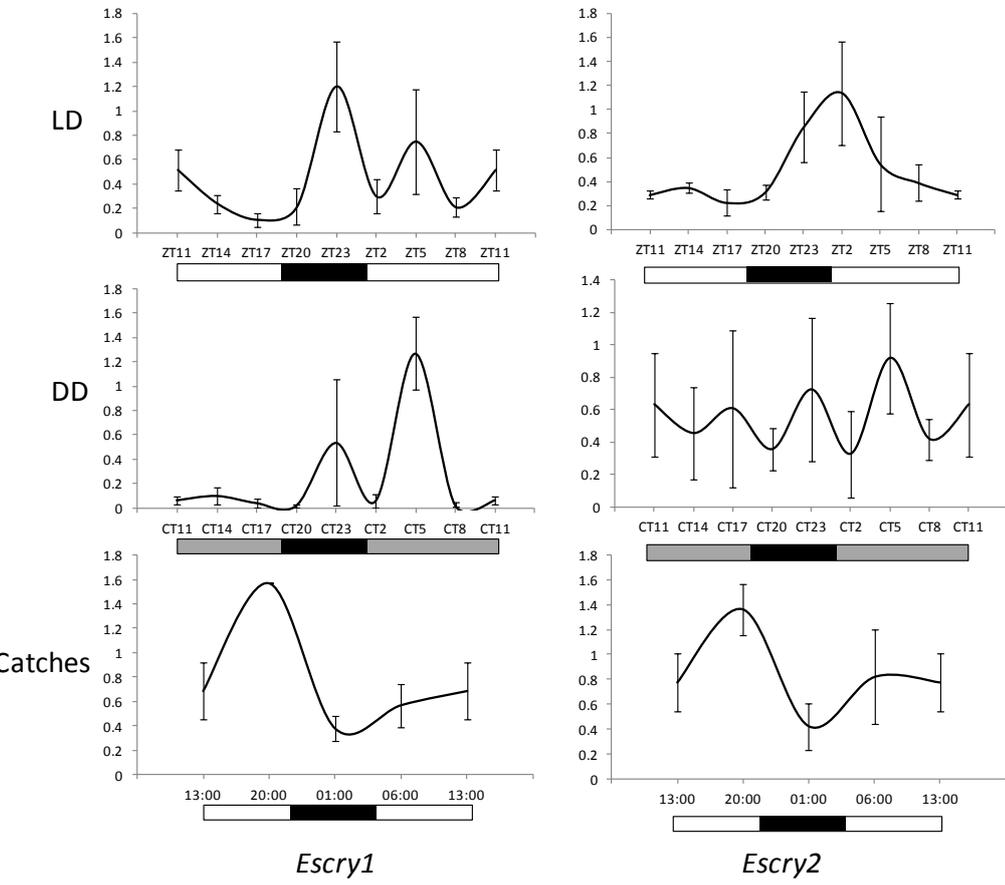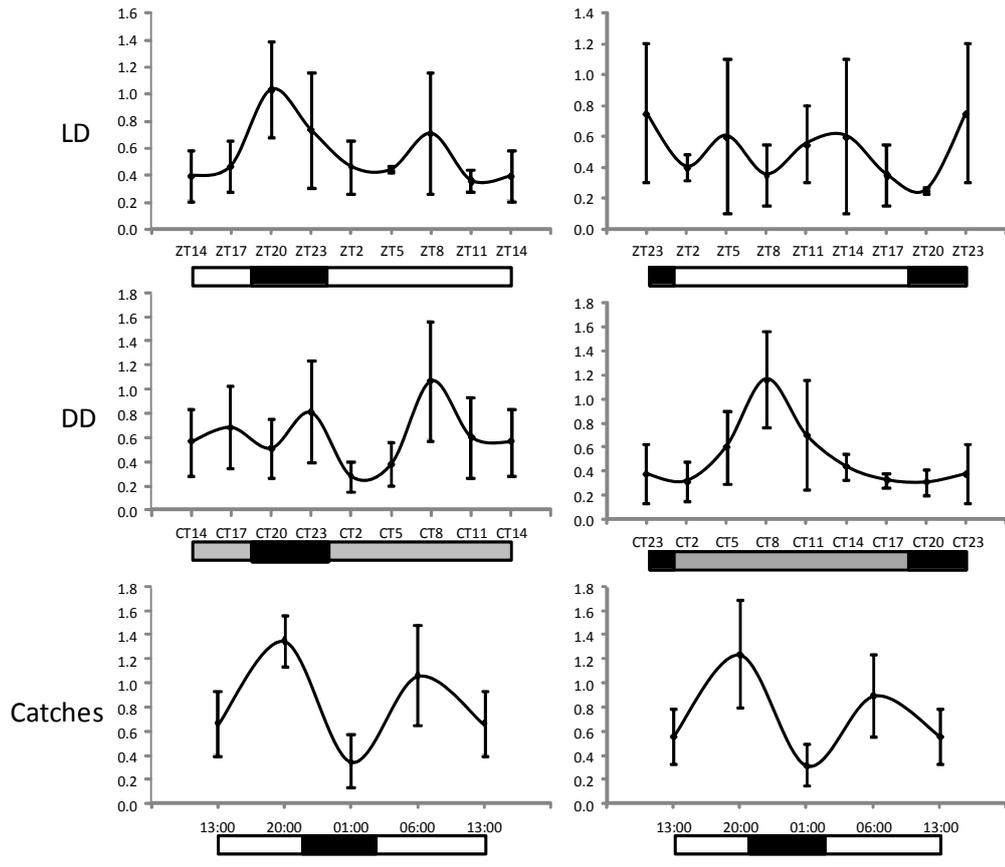
Given the contradictory results and the discrepancies in approach, it was considered that in order to confirm the current results and generate well-supported data for *Esper* and *Estim* RT-PCR should be repeated with the addition of housekeeping or reference gene(s) to increase assay robustness, and to this end a number of candidates have been identified and cloned through reference to the *Euphausia superba* transcriptome database (Chapter Three). Unfortunately it was not possible to conduct the necessary work during the time available before submission of this thesis, and this experiment remains the first point of order in any future research.

On the assumption that further work supports the initial results, a lack of transcript cycling would not be totally unexpected – the extreme changes in photoperiod experienced by *Euphausia* across the seasons may well encourage the evolution of an unconventional clock. Polar animals have been found to become behaviourally arrhythmic at certain times of year (van Oort *et al.*, 2005) and that could be reflected at the molecular level. That is the conclusion of a recent study on the Antarctic midge *Belgica antarctica*, which found no evidence of cycling of *period, timeless, Clock,* or the regulatory gene *vrille* in head tissue from animals in both field and lab (constant temperature and photoperiod) conditions (Kobelkova *et al.*, 2015). The authors raise the possibility of protein cycling in the absence of transcript cycling, however, something that should also be fully addressed in *Euphausia* should repeated RT-PCR fail to identify the latter.

2.4.4 Future work

2.4.4.1 Immediate issues

As stated above, a more complete picture of the molecular clock of *Euphausia* must be obtained through the generation of data detailing the temporal expression of all canonical clock genes, ideally across multiple days of DD. Assessment of protein cycling of all components will also be key to understanding the process.

Further cell work will prove illuminating. For example through deletion of particular regions such as the BCTR of EsBMAL1, the poly-Q of EsCLK and the putative PIS of EsTIM a more detailed model can be built of how the proteins interact to drive and repress gene expression. Transfected S2 cells subject to light pulses may ascertain if

EsCRY1 is capable of mediating EsTIM degradation as their orthologs do in *Drosophila* (Busza *et al.*, 2004). Nuclear localisation or otherwise of epitope-tagged repressive elements can be confirmed by immunofluorescence, to determine if EsPER and EsTIM are cytoplasmic when transfected individually and enter the nucleus when cotransfected. Coimmunoprecipitation experiments could be performed to confirm physical interactions, and again domain deletions could be employed to identify interaction sites, in line with the work of Kiyohara *et al.* (2006) that identified the BCTR as the site of mCRY repressive binding. It will be of fundamental importance to determine if the separate repressive limbs identified here represent a true picture of how *Euphausia*'s clock works, or if this phenomenon is an *in vitro* artefact enabled by an inherent capacity of the components to act as such, but that does not occur in the *in vivo* system.

2.4.4.2 Following up on a theory

In section 2.4.2.1 the suggestion was raised that the poly-Q region seen in many CLOCK proteins might be linked to local adaptation, possibly relating to the length of endogenous periods down a latitudinal cline or to wider seasonal responses. If one wished to follow up on this idea in krill, *Euphausia superba* may not be the ideal candidate – recent analysis suggests the entire circumpolar population to be panmictic and genetically homogeneous with little local adaptation (Deagle *et al.*, 2015). Instead we might turn to study of the Northern krill *Meganyctiphanes norvegica*, which occurs in waters ranging from the Mediterranean to the sub-Arctic (Tarling *et al.*, 2010) and has been shown to exist in genetically distinct populations (Zane *et al.*, 2000; Papetti *et al.*, 2005). Conditions in the Greenland Sea are undoubtedly different to those experienced in the Mediterranean, which raises interesting questions about the adaptability of the Euphausiid clock, and the *Clock* gene may be an excellent candidate with which to investigate the possible existence of latitudinal clines underlying local adaptation. We have successfully cloned a 792 bp fragment of *Meganyctiphanes Clk* using *Euphausia*-specific primers (Figure 2.16), providing a basis for further investigation.

```
MnCLK        1  FLSNEEFTHLMLEALDGFIMTISCSGRVLYTSESITPLLGHLPGDLSETP    50
                |||||||||||||||||||||:||||||||||||||||||||||||||:.
EsCLk        1  FLSNEEFTHLMLEALDGFIMTVSCSGRVLYTSESITPLLGHLPGDLSESP    50

MnCLK       51  LYDLMLNDERTDMRQFLSNPALAPDPTMRIDETKEQYAIAVHLKRAPTNL   100
                :|||||::||.|||:||||||||||:|::..:|:|||||||:||||||:||||:
EsCLk       51  VYDLMLSEERNDMRRFLSNPALAPNPSINMDDTKEQYAISVHLKRSPTNI   100

MnCLK      101  SEEPSYERVKLTGYFERYSCPSEDGVLDFSCSEAEDSVSVASSMVSSRQN   150
                |||||||||||||||||:|||||||||||||||:||..||..:||
EsCLk      101  SEEPSYERVKLTGYFERYTCPSEDGVLDFSCSEAEDSMSVDRSMFINRQ-   149

MnCLK      151  NNNNLGGMFGMTPSQSGVQDTKLVFVAIGRMERPQLVREMMILEPSKTEP   200
                |.||||:||.:||||||||||||||||||||||||||||||||||:||||
EsCLk      150  -NTNLGGIFGDSPSQSGVQDTKLVFVAIGRMERPQLVREMMILEPTKTEP   198

MnCLK      201  TSRHSLEWKFLFLDHRAPTIIGYLPFEVLGTSGYDYYHVDDLEKVSNCHE   250
                ||||||||||||||||||||||||||||||||||||||||:|||||||
EsCLk      199  TSRHSLEWKFLFLDHRAPTIIGYLPFEVLGTSGYDYYHVDDLDKVSNCHE   248

MnCLK      251  RLKKTGKGTSCYYR    264
                ||.|||||||||||
EsCLk      249  RIMKTGKGTSCYYR    262
```

Figure 2.16: Alignment of *Meganyctiphanes norvegica* CLK fragment with the equivalent *Euphausia superba* sequence. Red highlight = PAS domain.

2.4.5 Conclusions

A set of core circadian genes has been cloned and characterised for the Antarctic krill *Euphausia superba*, notably encompassing all key components of both the *Drosophila melanogaster* and *Mus musculus* negative feedback loops. Tissue specific PCR has confirmed expression of all components in the antennae, eye, head, thorax and abdomen. The key features of each peptide have been identified to draw inferences regarding their function and the wider evolution of the clock, with particular aspects of *Euphausia's* system appearing to represent an ancestral style. Luciferase reporter assays conducted using the *Drosophila* S2 cell system suggest that the clock has the ability to completely repress the transactivation activity of its central EsBMAL1:EsCLK heterodimer via two separate pathways.

Further work is required to characterise the temporal expression of the clock genes and their protein products, and explore the necessity and/or function of particular peptide domains. The apparent capacity of EsCRY2 and the combination of EsPER and EsTIM to both repress transactivation activity raises intriguing possibilities regarding their implementation *in vivo*, including seasonal responses and a model for the multiple oscillator hypothesis of crustacean chronobiology. Finally, the extensive poly-

Q domain seen in EsCLK is suggested as a starting point for further research into circadian latitudinal clines in krill species.

Chapter 3 The *Euphausia superba* transcriptome database: *de novo* assembly, annotation and output.

3.1 Introduction

3.1.1 Beyond the core

A transcription-translation negative feedback loop with a light sensitive component produces an entrainable system of rhythmic output – the basic molecular circadian clock is as neat a model as one might hope to see. As ever in biology it is rather more complex than that, and the wider picture remains poorly understood in certain aspects. A number of regulatory components work to enable and maintain circadian periodicity, without which the cycle would complete within a few hours (Gallego and Virshup, 2007). The activity and interactions of these components and the core system might be termed the 'complete' biological clock.

The details of these vary between the vertebrate and invertebrate model organisms studied. While intuition might suggest the full biological clock of *Euphausia superba* is likely to more closely resemble that of the invertebrate *Drosophila melanogaster* than the mammal *Mus musculus* it is possible that the crustacean clock could combine elements of both systems, considering the findings of the previous chapter. On that basis, in the following overview regulatory genes linked to the clock of both phyla will be covered.

3.1.1.1 The second loop

In *Drosophila,* the core transcription-translation loop (TTL) is lent further stability through the actions of a second loop, less well understood than the first. Two E-box genes, *vrille* and *Par domain protein 1 ε* are expressed through the actions of the CLK:CYC heterodimer (McDonald and Rosbash, 2001) and act as a repressor and activator of *Clk* respectively, causing transcript levels to oscillate.

Overexpression of *vrille*, which encodes a basic leucine zipper protein, has been shown to causing arrhythmicity in *Drosophila*, while heterozygous *vrille* mutant flies exhibit a shortened locomotor rhythm (Blau and Young, 1999). VRILLE acts through binding to

the V/P box (defined by the sequence TTATGTAA) present in the promoter region of *Clk* (Cyran *et al.*, 2003; Glossop *et al.*, 2003).

While expression of both genes is driven by the actions of CLK simultaneously and their mRNA transcripts accumulate at the same time, accumulation of the activator protein PDP1ε - and thus its contribution to the second loop - lags behind that of VRILLE by 3 to 6 hours. Homozygous *pdp1ε* mutant flies are arrhythmic and show much reduced expression of *Clk*, while heterozygotes show a lengthened behavioural rhythm (Cyran *et al.*, 2003). The role of PDP1ε in the molecular clock is controversial, having been suggested to instead act on behavioural output (Benito *et al.*, 2007) or at least be just one of a number of components acting to drive *Clk* expression (Zheng *et al.*, 2009).

In mammals, the second loop involves control of expression of the other half of the heterodimer, BMAL1. Two orphan nuclear receptors, Rora and Rev-Erbα, expressed by the activity of the core heterodimer, play the roles of repressor and activator respectively (Sato *et al.*, 2004; Preitner *et al.*, 2002), generating oscillations in Bmal1 transcript levels.

3.1.1.2 Transcriptional regulators

Regulation of the expression of PER and TIM has been linked to NEJIRE (NEJ), a transcription factor whose role is currently unclear. Lim et al. (2007) found that overexpression resulted in lower levels of PER *in vivo* and further work using embryonic *Drosophila* S2 cells showed it to repress activity of the CLK:CYC heterodimer. Conversely Hung et al. (2007), again working with S2 cells, found that down-regulation of NEJ activity resulted in decreased CLK:CYC activity, and the status of NEJ as a transcriptional activator or repressor remains uncertain.

A second transcriptional regulator is the basic helix-loop-helix-ORANGE protein CLOCKWORK ORANGE (CWO). Again its precise role is unclear, with evidence suggesting both activator and repressor roles including the ability to directly repress its own transcription (Richier *et al.*, 2008). S2 cell assays have shown CWO's ability to repress promoters containing E-box sequences (Matsumoto *et al.*, 2007), yet robust oscillation of *per*, *tim*, *vri* and *Pdp1ε* mRNA is lost in the *cwo* null mutant through a failure to drive their evening expression (Richier *et al.*, 2008).

3.1.1.3 Post-translational regulation

Abolishing PER and TIM protein cycling destroys rhythmicity, but the same is not true of their mRNA cycling (Yang and Sehgal, 2001) and it now thought that post-translational modifications of the clock proteins by enzymes such as kinases, E3-ubiquitin ligases and phosphatases may be key in maintaining rhythmicity (Özkaya and Rosato, 2012).

PERIOD is a protein subject to much phosphorylation. A key step in producing the delay required to generate a 24 hour rhythm is the degradation of PER, thus slowing down accumulation of the protein and its subsequent entry into and repressive activity within the nucleus. The crucial stage is the phosphorylation of the residue Ser47 by the kinase DOUBLE-TIME (DBT; Kloss et al. 1998, Price et al. 1998), promoting its degradation via the E3-ubiquitin ligase SUPERNUMARY-LIMBS (SLIMB; Chiu et al. 2008). The phosphorylation of Ser47 is itself delayed by the work of another kinase, NEMO (NMO; Chiu et al. 2011), which promotes DBT phosphorylation activity at other residues. Phosphorylation of Ser657 by SHAGGY (SGG; Ko et al. 2010), an ortholog of the mammalian kinase GSK-3β, meanwhile, promotes the shuttling of PER from the cytoplasm to the nucleus.

SHAGGY also interacts with TIM to direct entry to the nucleus; overexpression leads to increased and premature translocation and a shortened locomotor rhythm as a result, while *sgg* mutants show decreased levels of TIM phosphorylation (Martinek *et al.*, 2001). Casein kinase II (CKII) is another kinase that phosphorylates both PER and TIM. The kinase is a tetramer built from two subunits, CKIIα (catalytic) and CKIIβ (regulatory), with mutations in both shown to alter locomotor rhythms and lower PER and TIM degradation (Akten *et al.*, 2003; Meissner *et al.*, 2008).

Protein phosphatase I (PPI) and Protein phosphatase II (PPII) are further contributors to post-translational control of the clock. Reduced expression of the former decreases levels of TIM, delays its nuclear accumulation and both lengthens and weakens the locomotor period, possibly through interactions with SGG (Fang *et al.*, 2007). The latter, a heterotrimer composed of the subunits MICROTUBULE-STAR (MTS), TWINS (TWS) and WIDERBORST (WDB), appears to interact with PER. Over- or

underexpression of certain subunits have been shown to respectively shorten or lengthen locomotor activity periods (Sathyanarayanan *et al.*, 2004).

CIRCADIAN TRIP (CTRIP) is, like SLIMB, an E3-ubiquitin ligase implicated in degradation of PER (Lamaze *et al.*, 2011). Downregulation through RNAi lengthens the locomotor activity period and results in high levels of PER and CLK, suggesting it is also involved in turnover of the latter.

Indeed in *Drosophila,* while CYC is constitutively expressed, CLK is itself subject to post-translational regulation to generate protein activity cycling across 24 hours. Transcription is prevented by the phosphorylation of CLK through the actions of PER and TIM bound to DBT (Kim *et al.*, 2007) and possibly NEMO (Yu *et al.*, 2011).

3.1.1.4 Light mediated interactions

Chapter Two describes the role of light-activated CRY and the subsequent degradation of TIM in the *Drosophila* clock. The F-box protein JETLAG (JET) is an intermediary in this process, binding to bound CRY:TIM complexes in order to initiate degradation (Peschel *et al.*, 2009). Degradation of TIM is lessened in mutant *jet* flies, which also fail to exhibit arrhythmia under constant light (Koh, 2006).

3.1.1.5 Clock controlled genes

The following are a small selection of genes specifically identified as under the control of the circadian clock.

LARK is an RNA-binding protein that shows rhythmic protein abundance even in constant conditions, but which is abolished in *per* mutant flies (McNeil *et al.*, 1998). The protein is thought to play a role in regulation of eclosion, a well-established circadian assay in *Drosophila* (Skopik and Pittendrigh, 1967). TAKEOUT (TO) is a protein induced through starvation; expression is localised to tissues involved in feeding and *to* mutants succumb to starvation rapidly. It shows mRNA cycling that is reduced or abolished in circadian mutant flies (Sarov-Blat *et al.*, 2000).

The classic clock controlled gene of *Drosophila* is the neuropeptide *Pigment-dispersing factor* (*Pdf*). PDF is expressed in the lateral ventral neurons (LN$_V$), key to control of circadian behaviour (Helfrich-Förster *et al.*, 1998), and *Pdf* mutants show abolishment

of rhythmic behaviour in constant conditions, as do flies with ablated *Pdf* neurons (Renn *et al.*, 1999). Flies mutant for *Clk* and *cyc* show greatly reduced levels of mRNA and PDF peptide (Park *et al.*, 2000). In crustaceans the ortholog is termed *Pdh* (*Pigment dispersing hormone*) and several genes encoding PDH peptides have been identified in various species, subdivided into α and β forms (Rao, 2001). The latter in the crab *Cancer productus* has been shown to serve as a functional replacement for PDF in *pdf*01 *Drosophila* mutants (Beckwith *et al.*, 2011) and PDH has been linked to the circadian clock in other crustaceans (Verde *et al.*, 2007). In *Euphausia superba* two forms, PDH-Lα and PDH-β, have been identified (Toullec *et al.*, 2013),

In mammals, a key hormone controlled by the circadian clock is melatonin (Cassone, 1990). This has been shown to cycle in the cladoceran *Daphnia pulex* via oscillating expression of insect-like arylalkylamine N-transferases (AANATs), the enzyme driving melatonin synthesis (Schwarzenberger and Wacker, 2015). It has also been linked to diel vertical migration - the behaviour of *Euphausia* identified as a candidate for circadian regulation - in *Daphnia magna* (Bentkowski *et al.*, 2010) and the annelid *Platynereis dumerilii* (Tosches *et al.*, 2014).

3.1.1.6 Other neuropeptides and hormones

Certain other crustacean neuropeptides and hormones have been implicated in circadian output. Red pigment concentrating hormone (RPCH), an antagonist of PDH, shows circadian protein and mRNA cycling in crab and crayfish that persist in constant conditions (Fingerman and Fingerman, 1977; Sosa *et al.*, 1994) and is considered an important molecule in regulating the circadian rhythm of the eye. Crustacean hyperglycaemic hormone (CHH), linked to carbohydrate metabolism and blood sugar regulation, follows a circadian pattern of abundance (Kallen *et al.*, 1990).

Drawing the net still wider, while direct evidence of involvement may be lacking in many cases, it can be argued that many other signalling molecules are likely tied to the output of the biological clock, being so deeply enmeshed in physiology, metabolism and behaviour. Christie *et al* (2010b) provide a comprehensive overview of crustacean neuropeptides that might prove to be controlled by the clock.

3.1.2 High throughput sequencing

3.1.2.1 So many genes, so little time

Chapter Two describes the cloning of the core circadian genes through the use of degenerate PCR or by mining NCBI databases for fragments, with subsequent RACE extension used to complete the coding sequence. The above introduction sets out the wider molecular picture of the *Drosophila* and *Mus* clocks; though decades of research underpins such extensive detail a researcher hoping to accurately depict the clock of another species should consider the regulatory and output components as well. Degenerate PCR, however, is not a technique particularly suited to identifying large numbers of genes. Often it is difficult to identify suitably conserved regions that offer the opportunity to design suitable degenerate primers, or such sites are so widely spaced the PCR efficiency is affected. The mixture of nucleotides at particular positions used in degenerate primers raises the risk of amplification of sequences other than the target, and it can be a slow and painstaking process. Unsuccessful efforts were made to clone *Estimeless*, *Esperiod* and *Esvrille* using this technique before the EST database method was attempted.

Similarly, RACE extension is not always straightforward and often requires optimisation and repeated attempts to obtain the full 5' and 3' sequences as reactions can produce truncated fragments. These are time-consuming techniques most suited to cloning only a few genes of interest at most; to identify a full suite of circadian-related genes in this way is simply not a sensible use of time and resources.

3.1.2.2 Genome mining

Gene identification is much quicker and easier if the sequencing has already been completed. If an organism has a sequenced genome then known genes from other species can be used to identify orthologs in the species of interest using an alignment search tool. This technique was used to identify a circadian system in *Daphnia pulex* (Tilden *et al.*, 2011), at the time only the third crustacean species in which circadian genes had been identified. Using protein sequences of the core, regulatory and output clock genes of *Drosophila* as queries, the group identified and characterised 30

predicted protein sequences including the core components CLOCK, CYCLE, PERIOD, TIMELESS, CRY1 and CRY2.

This exact method is not currently an option for *Euphausia,* which lacks a sequenced genome; this situation is unlikely to change soon with the genome size estimated at 47 Gbp (Jeffery, 2012) – incidentally a finding that appears to be something of a trend in polar crustaceans (Rees *et al.,* 2007; Rees *et al.,* 2008). In recent years however a sequencing technology has emerged that enables a very similar process to be undertaken at a fraction of the time and cost.

3.1.2.3 RNA-seq

RNA-seq uses high-throughput sequencing technology on cDNA fragments created from RNA transcripts to characterise the transcriptome, a complete example of which would represent all transcripts present in the cell under the specific conditions at the point of RNA extraction.

The sequencer (platforms available include those made by Illumina, Roche and Applied Biosystems) generates millions of short reads, typically from 50 to 400 bp in length. Read can be single-end or paired-end, the latter generating reads from both ends of a fragment. This provides extra information for subsequent analysis in that the aligner or assembler knows that the paired reads come from the same transcript even if they do not overlap, allowing higher quality alignments and scaffolding of unconnected contigs. Once sequenced, reads can be used for a number of downstream analyses, such as alignment to a reference genome to measure transcript abundance; identification of differential expression between samples under different conditions; or to identify isoforms.

One of the major appeals of RNA-seq for those who work with non-model organisms lacking a reference genome is that the reads can be assembled into a *de novo* transcriptome assembly. This has been conducted for many species in recent years, including *Euphausia superba*, albeit on a relatively small scale using hundreds of thousands of reads rather than millions (Clark *et al.,* 2011; De Pittà *et al.,* 2013; Martins *et al.,* 2015). As with the genome of *Daphnia pulex*, a transcriptome assembly

for the copepod *Calanus finmarchicus* has been mined for the putative components of its circadian clock (Christie *et al.* 2013a).

3.1.2.4 Building a *de novo* transcriptome assembly

An assembly is collection of contigs, each one ideally a complete reassembly of an RNA transcript, and can serve as a proxy genome to which reads can be mapped to achieve the kind of analyses listed above, as well as enable the discovery of thousands of previously unknown coding sequences. There are a number of non-commercial assemblers available, many of them using the de Brujin graph method, more suited to short read assembly and pioneered by Pevzner *et al*. (2001) with the EULER assembler; before this time the 'overlap-layout-consensus' approach was dominant.

The de Brujin graph method works by breaking the reads down into even smaller fragments of length *k*, called *k*-mers. These are designated as nodes and then connected provided they overlap by *k*-1. Branches or bubbles in the graph are created as SNPs, sequencing errors, or apparent deletions or introns are detected, and then the graph is collapsed into larger nodes where branching is not present and transcript paths output as individual contigs (Martin and Wang (2011); Figure 3.1).
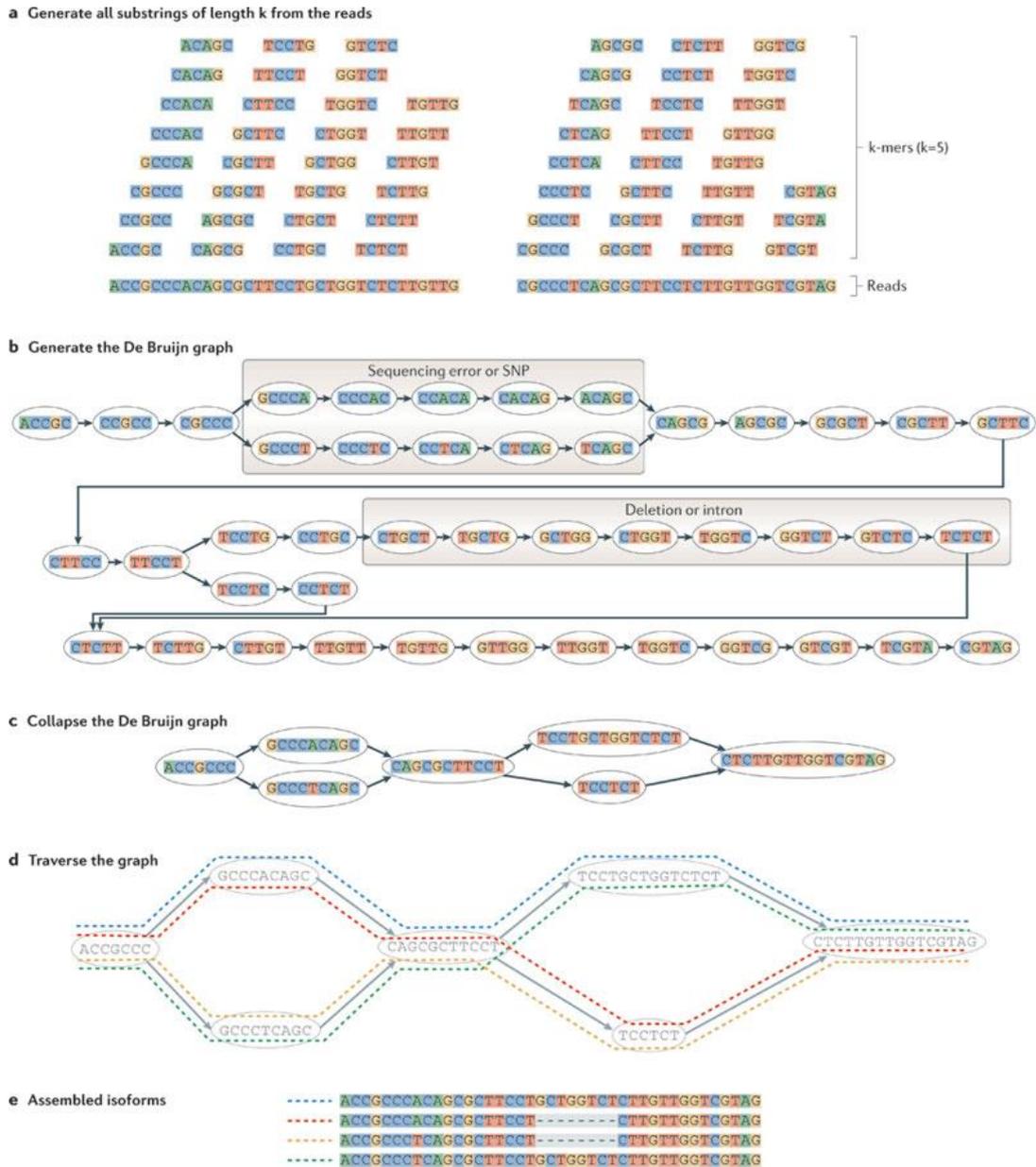
Figure 3.1: De Brujin graph assembly strategy. Reads are broken into substrings of length *k* and linked as nodes in the graph, with bubbles or branches deriving from single nucleotide differences or deletions/introns. Nodes are collapsed and the graph traversed, resulting in assembly of all possible isoforms. From Martin and Wang (2011).

Perhaps the most important step in optimising the assembly of a *de novo* transcriptome involves exploiting the range of *k*-mer values used. This parameter can have a great effect on the final assembly (Robertson *et al.*, 2010). Small *k*-mers perform better at assembling lowly expressed transcripts and favour transcript diversity but at the cost of accuracy and fragmentation, while large *k*-mers have the

advantage of high accuracy but produce fewer, though longer, transcripts from highly expressed genes (Surget-Groba and Montoya-Burgos, 2010). The popular assembler Trinity (Grabherr *et al.*, 2011) forces use of an intermediate *k*-mer of 25 as a compromise, but other assemblers allow the user to define the *k*-mer parameter when assembling. The optimal approach is likely to employ the latter to create multiple assemblies across the range of possible *k*-mers that are then merged into a single assembly and processed to remove redundancy, with tests showing that each *k*-mer assembly is capable of producing unique contigs (Haznedaroglu *et al.*, 2012). While many assemblers share the de Brujin graph method as their basis, they too are likely to construct contigs unique to their particular algorithms. It is reasonable to assume that if one wishes to maximise the number of unique, true contigs in order to facilitate gene discovery, a multi-assembler, multi-*k*-mer approach is likely to prove productive.

De novo transcriptome assembly is an error-prone process. The final assembly inevitably comprises correctly assembled contigs - some fragmented or redundant - plus those generated from the collapse of gene families, chimeras and other misassemblies, (Figure 3.2).
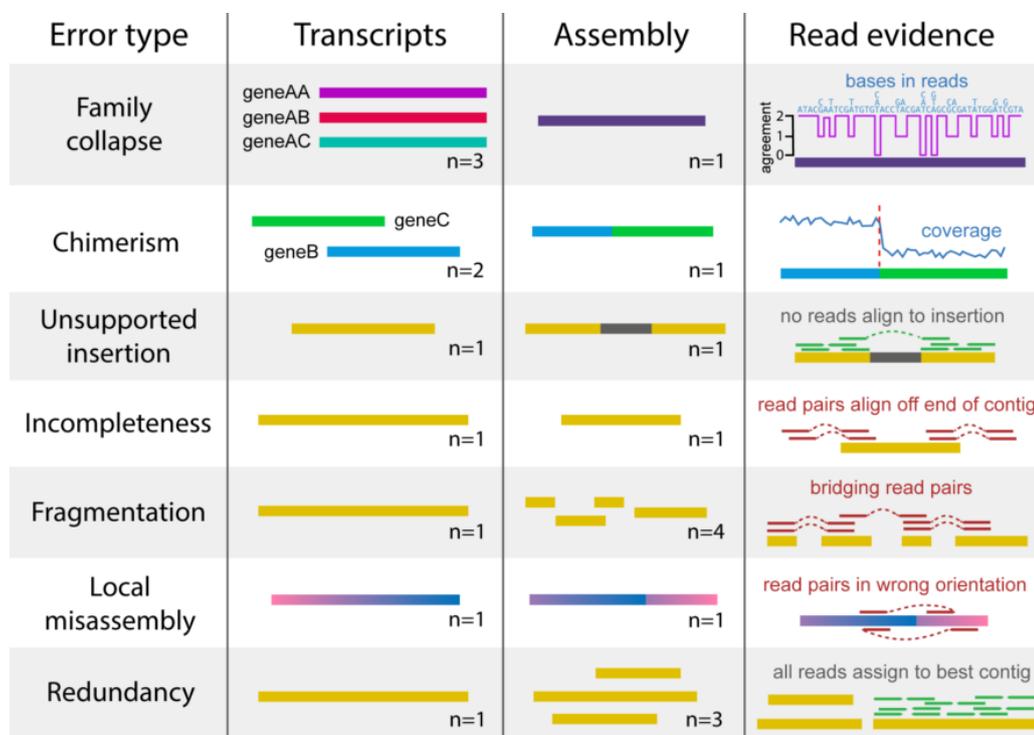
Figure 3.2: Types of errors inherent in the assembly of *de novo* transcriptome contigs. From Smith-Unna *et al* (2015).

A transcriptome produced by merging many separate assemblies from different pieces of software using different *k*-mer parameters will be particularly vulnerable to such problems; unique misassemblies will accumulate while only small numbers of unique, genuine contigs may be added in the contribution of each assembly. To some extent it is the price to pay for attempting to maximise the potential for gene discovery, but a number of programs have been developed to improve the quality assembly quality. These include tools to reduce redundancy at a user-defined similarity threshold (Li and Godzik, 2006; Fu *et al.*, 2012), tools to predict which contigs contain open reading frames and are likely to code for putative peptides (Haas *et al.*, 2013), and tools that assess the quality of a contig via read mapping analysis such as Transrate (Smith-Unna *et al.*, 2015). This latter in particular is designed to identify those issues identified in Figure 3.2, assigning high contig scores to those with the best quality read evidence. The Transrate contig score is based on the proportion of nucleotides in a contig with read mapping support, the extent to which the bases of mapped reads match those of the contig, the extent to which bases are correctly ordered as determined by read pair data, and the degree of coverage consistency across the contig (chimeras are likely to show inconsistency). Further to the use of such tools, subsequent annotation of the transcriptome will identify by homology and identification of domains and motifs which sequences are of high quality.

The rest of this chapter describes the assembly of a new, comprehensive *de novo* assembled transcriptome for *Euphausia superba,* its annotation and development into an accessible online resource for researchers, and its exploitation in deriving a full suite of circadian-related coding sequences.

3.2 Methods

3.2.1 Assembling the transcriptome

3.2.1.1 Illumina sequencing

Total RNA was precipitated from each catches timepoint sample (see section 2.2.1.1), combined into a single sample of quantity > 10 µg and concentration > 250 ng/µl and delivered to BGI Tech Solutions (Hong Kong) Co Ltd for library construction and paired end sequencing using the Illumina HiSeq 2000 platform. The RNA sample was subject to polyA mRNA enrichment using oligo(dT) beads prior to fragmentation and cDNA synthesis. Adaptor sequences, contamination and low quality reads were removed from the raw data which was then retrieved via FTP in the form of two FASTQ files. Reads were subject to quality control analysis once downloaded using FastQC v0.11.2 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and low quality (≤ 3) leading and trailing bases were identified and removed using Trimmomatic 0.32 (Bolger *et al.*, 2014).

3.2.1.2 Assembly

A multi-assembler, multi-*k*-mer approach was adopted. Assemblies were created using SOAPdenovo-Trans (Xie *et al.*, 2014), Trans-ABySS (Robertson *et al.*, 2010), Bridger (Chang *et al.*, 2015) and Trinity (Grabherr *et al.*, 2011). For both SOAPdenovo-Trans and Trans-ABySS eight separate assemblies were created using *k*-mers ranging from 21 to 91 in 10 step increments. Bridger has a maximum *k*-mer setting of 31; six separate assemblies were therefore built from 21 to 31 in two step increments. Trinity does not allow the user to choose a *k*-mer value so a single assembly was built using this software with the fixed *k*-mer value of 25. A minimum contig length of 200 was specified where possible – where it was not, contigs shorter than 200 were removed post-assembly.

The output of each assembler that generated multiple assemblies was merged into assembler-specific files. Trans-ABySS has a built-in merge function for this purpose. For SOAPdenovo-Trans the contig headers in each assembly were manually amended with a prefix to reflect the *k*-mer used and then the complete assemblies were combined with a simple concatenation command. Redundancy was addressed using the

dedupe.sh function of BBMap (http://sourceforge.net/projects/bbmap/), the parameters set to remove duplicates at 100% identity and containments (shorter contigs contained perfectly within longer ones) including those registering only when read in reverse complement. The same process was used to merge the Bridger assemblies.

3.2.1.3 Quality assessment of assemblies and contigs

Each merged, deduplicated assembler-specific file was subject to quality assessment using Transrate 1.0.1 (Smith-Unna *et al.*, 2015). The best quality contigs from each, as determined by the Transrate contig score criteria, were combined into a single file and subject to de-duplication at 100% identity using BBMap. Finally, Transrate was run on this file to select the absolute best contigs among all assemblers. Going forward, this selection of contigs will be referred to as the **total assembly**.

3.2.2   Annotation and analysis

3.2.2.1 Identification of coding contigs

The total assembly was subject to processing using Transdecoder (Haas *et al.*, 2013), a piece of software written to identify coding regions within transcripts. Default parameters were used: minimum open reading frame (ORF) coding for a peptide length of 100 amino acids. Transdecoder produces two peptide files, one simply listing all ORFs coding for a peptide of 100 aa or more, the second uses the longest ORFs as parameters in a Markov model to assign coding scores to each ORF and reports what it determines to be likely coding peptides. The latter file was processed using CD-HIT (Li and Godzik, 2006; Fu *et al.*, 2012) at 100% identity to remove duplicates at the protein level that would have not been considered duplicate at the nucleotide level due to the redundancy of the genetic code. Going forward, the resulting file will be referred to as the **peptide assembly**. The contigs that coded for these peptides were extracted from the total assembly: this selection of contigs will be referred to as the **coding assembly**.

3.2.2.2 Annotation by homology

The peptide assembly was queried against a BLAST protein database constructed from the arthropod dataset of the UniProt Knowledgebase (ftp://ftp.ebi.ac.uk/pub/databases/fastafiles/uniprot/, retrieved 4th August 2015) using

the blastp function of the command line BLAST+ software (Camacho *et al.*, 2009). Parameters were set to ensure each query returned at most a single result with an E-value of $1.0e^{-6}$ or lower. The output was set to XML format with GenInfo Identifiers (GI) shown to ensure compatibility with subsequent annotation steps.

3.2.2.3 GO terms

The peptide assembly and BLAST XML output file were loaded into BLAST2GO (Conesa *et al.*, 2005), an annotation toolset. As the BLAST step had already been completed using the much quicker local database process described above, this was skipped. The remaining steps were completed with default settings.

3.2.2.4 The *Euphausia superba* transcriptome database

For use in the *Euphausia superba* Transcriptome Database (EsT-DB), the data were processed and converted into an appropriate format. For the searchable front end both the peptide assembly and coding assembly were merged with the annotation data and then given sequential and consistently-formatted gene, transcript, and ORF IDs as well as being assigned putative gene names based on the BLASTP results. The EsT-DB was created using a Catalyst framework to access annotated transcriptome data stored in a MySQL database, and the front-end design makes use of template toolkit and Twitter Bootstrap. As an extra resource the peptide assembly, coding assembly and total assembly were made available as BLAST databases using the SequenceServer front end (Priyam *et al.*, 2015).

3.2.3   Mining the transcriptome

All queries were conducted on the total assembly, on the logic that even very truncated coding sequences could prove vital in identifying a complete transcript.

3.2.3.1 Core circadian genes

The sequences of the core circadian genes already fully or partially cloned (see previous chapter) were used to query the transcriptome due to the possible existence of isoforms and/or to attempt to complete the partially sequenced genes *Esperiod* and *Estimeless*. These sequences were also used as queries against a BLAST database created using the single Trinity assembly as a basic assessment of the completeness of the total assembly against the single *k*-mer output of one piece of software.

3.2.3.2 Regulatory and output genes

Protein sequences derived from genes from *D. melanogaster* or *M. musculus* known to contribute to or interact with the circadian clock were used to search for *Euphausia* orthologs using the tblastn function of BLAST+ with an E-value cutoff of $1.0e^{-3}$. Transcriptome contigs returned by this search were extracted from the database, translated and used as input queries in a reverse-BLAST process to search the NCBI non-redundant (NR) protein database (excluding uncultured/environmental sample sequences and all entries with a title containing the words '*putative*', '*hypothetical*' or '*predicted*') and Flybase *D. melanogaster* database (version FB2014_03) for the proteins most similar to each. These sequences were accepted or rejected as putative orthologs of known circadian genes of the basis of this reverse-BLAST analysis. Where circumstances indicated such actions might be fruitful, further searches of the transcriptome were conducted using query sequences from species more closely related to *Euphausia* – such instances are specified in the results.

Where the results suggested a full coding sequence for key putative regulatory proteins had been assembled specific primers were designed in order to reamplify the sequence via PCR using Q5 polymerase (NEB) per manufacturer's instructions. Amplicons were gel-extracted, cloned, sequenced and aligned with the original transcriptome contig. If an amplicon showed minor differences from the contig, the former was accepted as the correct sequence if confirmed with repeated independent sequencing. For certain genes of particular importance to the circadian system, attempts were made to complete fragmentary contigs through RACE extension.

Each output contig was aligned with its top BLAST hit from Flybase and NCBI NR database in order to determine amino acid identity and similarity. Pairwise alignments were generated using EMBOSS Needle Pairwise Sequence Alignment (http://www.ebi.ac.uk/Tools/psa/emboss_needle/). The PDH-Lα multiple alignment was generated using Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/). Protein domains and structural motifs were identified using SMART (http://smart.embl-heidelberg.de/; Schultz *et al*. 1998, Letunic *et al*. 2015) with default parameters, and the NCBI Conserved Domain Search function.

Given the complexity and diversity of the GNAT superfamily to which insect-like arylalkylamine N-transferases belong, a neighbour-joining phylogenetic tree was generated with 500 bootstrap replications and complete deletion using MEGA 7.0 with MUSCLE alignment, to ascertain where orthologous *Euphausia* contigs might cluster.

3.2.3.3 Neuropeptides and receptors

A list of accession numbers to use as input queries in identifying potential *Euphausia* preprohormones and associated receptors was assembled from a series of relevant studies of various species (Gard *et al.*, 2009; Ma *et al.*, 2009; Christie *et al.*, 2010a; Christie *et al.*, 2010b; Ma *et al.*, 2010; Christie *et al.*, 2011a; Christie *et al.*, 2011b; Christie *et al.*, 2013b). The protein sequences were retrieved from the NCBI protein database and used to search for *Euphausia* orthologs using the tblastn function of BLAST+.

Transcriptome contigs returned by this search were extracted from the database, translated and subject to a reverse-BLAST analysis as before. Sequences were accepted or rejected as putative orthologs of known hormone/receptor genes of the basis of this reverse-BLAST analysis.

3.2.4   Transcript abundance

To estimate the abundance of transcripts the quality-trimmed reads were aligned to the 'good' contigs (as determined by Transrate) of the individual 25 *k*-mer Bridger assembly, selected as an intermediate *k*-mer assembly from the assembler with the highest overall N50 metric (see Table 3.1 A). Trinity's alignment plug-in *align_and_estimate_abundance.pl* was used, employing RSEM as the aligner. Contigs of interest in this particular assembly – which were not necessarily those identified in the process described in previous sections due to the combinatory and de-duplicated nature of the total assembly – were identified using the tblastn function of BLAST+, with the contigs from the total assembly used as input queries. As a number of the core circadian genes were fragmented a second abundance calculation was performed on the same transcriptome in which these fragments had been removed and replaced with the full coding sequences including 5' and 3' UTRs, in order to determine if the fragmentation had any notable effect on TPM (transcripts per million, a relative

measure of abundance; Wagner, Kin and Lynch (2012)). The affected genes were *Esbmal1*, *Esclock*, *Escryptochrome1* and *Estimeless*.

A further analysis was conducted to estimate how many transcripts were well supported by the expression data, following the procedure outlined at https://github.com/trinityrnaseq/trinityrnaseq/wiki/Trinity-Transcript-Quantification#counting-expressed-genes.

 The Trinity plug-ins *abundance_estimates_to_matrix.pl* and *count_matrix_features_given_MIN_TPM_threshold.pl* were used to generate a count of transcripts grouped by TPM. These data were filtered to remove transcripts with a TPM of less than 10 and greater than 100 and subject to a linear regression, subsequently used to extrapolate to an estimate of the number of expressed genes based on the trend between these two abundance levels.

3.3   Results

3.3.1   Transcriptome assembly and annotation

3.3.1.1 Sequencing data and QC

A total of 69,837,314 clean reads were generated by the Illumina sequencing, specifically 34,918,657 paired-end 100 base pair reads. The summary report provided by BGI Tech Solutions determined that 98.62% of sequenced bases were of Phred quality 20 or greater with a GC percentage of 44.06.
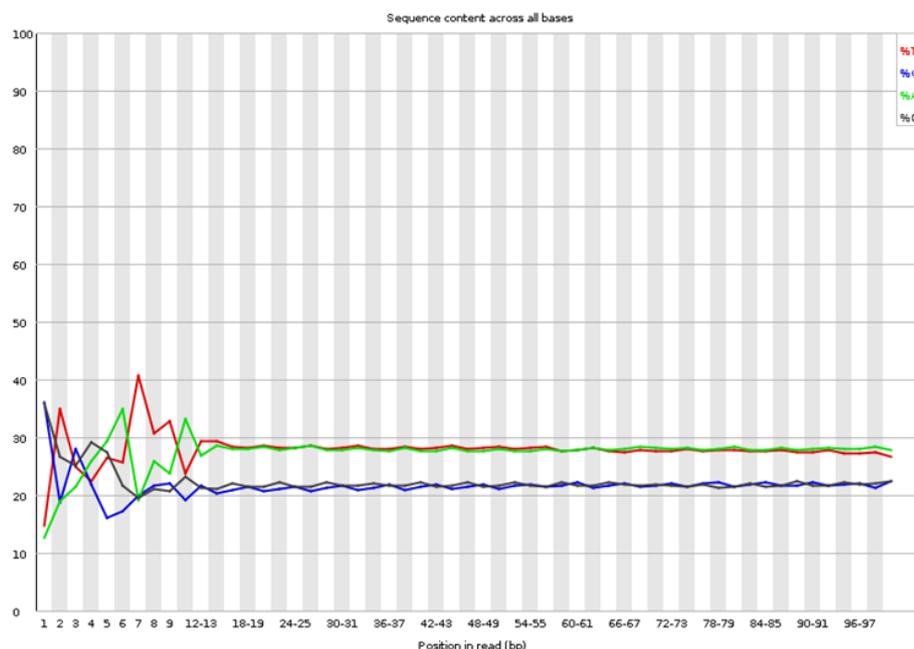
Figure 3.3: FastQC Report on sequence content across all bases for unprocessed read pair 1 sequencing file.

Analysis of the sequencing files using FastQC indicated that the first 10 – 13 bases of the reads had biases towards particular nucleotides (Figure 3.3) that is explained by the random hexamer priming methodology of Illumina sequencing (Hansen *et al.*, 2010) and would be unlikely to cause issues with assembly. Visual inspection of the sequencing files suggested a pattern of a single low quality undefined base at the beginning of a number of reads, which were removed by the Trimmomatic processing for low quality bases at each end of all reads. After processing the read lengths ranged between 81 – 100 bases.

3.3.1.2 Assembly assessment: individual and combined

Table 3.1 A shows the basic statistics for assemblies generated by each individual piece of assembly software, post-merge for those assemblers capable of multi-*k*-mer output. Bridger produced the most contigs by a clear margin, with Trinity producing only 30% of Bridger's output. The mean contig length, contig N50 score (50% of all bases in an assembly are in sequences of length N or greater) and number of contigs with an open reading frame was also greatest for the Bridger assembly. Trans-ABySS produced the longest contig of the group, while the SOAPdenovo-Trans assembly generated the lowest mean contig length and N50, and the lowest number of contigs with an open reading frame as a percentage of the complete assembly.

Table 3.1 B shows further assembly metrics as generated by Transrate. Detailed explanations for each metric are shown in Table 3.2. Good mappings are defined as those where both members of a read pair map to a contig in the right orientation without overlapping either end of that contig; a bad mapping fails at least one of these conditions. Coverage is calculated per base, a nucleotide defined as covered if at least one read maps to it; further metrics are used to assess the assembly based on this score (number of contigs with at least one uncovered base or number of contigs with a mean coverage < 1, for example). Bridges are potential links between contigs that could in theory be assembled into a larger transcript; evidence of a fragmented assembly. The coding assembly was not subject to read mapping metrics due to the exclusion of a large number of contigs, which would make the numbers meaningless.

Bridger again performed well in major metrics, producing an assembly that 97% of reads mapped to, and 88% of which were considered good mappings. However the number of bases uncovered by any reads was an order of magnitude higher in the Bridger assembly; other coverage metric scores are similarly poorer than those received by the other assemblers and the final Transrate score for Bridger was the poorest of the four. Trinity, meanwhile, mapped the lowest number of reads and had the lowest percentage of good mappings but show superior coverage metrics in comparison to the other three assemblers. Bridger showed the greatest potential for improvement by using only the good contigs as determined by Transrate's optimal assembly score, while Trinity showed the least.

3.3.1.3 Total, coding and peptide assemblies

Table 3.1 A and B shows the Transrate scores for the total assembly. Absolute improvements are seen in the increase in the number of reads mapped and number of good mappings, the decrease in bad mappings and potential bridges and the overall Transrate score. Coverage metrics show a bias towards those seen in the initial Bridger assessment, as do many of the basic statistics.

From the total assembly comprising 484,125 contigs Transdecoder identified 149,683 potential peptides of 100 amino acids in length or longer, saved as the peptide assembly. These peptides originated from 140,552 coding contigs that make up the coding assembly.

Table 3.1: Quality analysis of individual and combined assemblies generated with Transrate. A) contig metrics. B) read mapping metrics.

A	Metric	Assembler					
		Bridger	SOAPdenovo-Trans	Trans-ABySS	Trinity	Total Assembly	Coding assembly
	No. of contigs	513,499	282,833	318,436	157,361	484,125	140,522
	Shortest	189	200	200	201	189	297
	Longest	27,543	17,265	37,689	14,517	34,468	34,468
	Mean length	780	445	573	504	726	1,157
	No. with ORF	138,937	51,336	72,597	31,588	135,969	100,311
	ORFs %	27%	18%	23%	20%	28%	81%
	N50	1,218	480	738	605	1,072	1,575

B	Metric	Assembler				
		Bridger	SOAPdenovo-Trans	Trans-ABySS	Trinity	Total Assembly
	Reads	34,918,657	34,918,657	34,918,657	34,918,657	34,918,657
	Reads mapped	33,878,827	32,480,890	32,111,813	29,500,220	34,075,930
	% of reads mapped	97%	93%	92%	84%	98%
	Good mappings	30,880,238	28,292,720	28,688,483	22,359,263	31,833,986
	% good mappings	88%	81%	82%	64%	91%
	Bad mappings	2,998,589	4,188,170	3,423,330	7,140,957	2,241,944
	Potential bridges	43,048	46,708	38,398	37,910	34,120
	Bases uncovered by any reads	151,034,498	13,315,462	27,831,787	2,882,165	79,770,922
	% uncovered bases	38%	11%	15%	4%	23%
	Contigs with uncovered bases	441,180	199,695	245,441	68,662	388,395
	% contigs with uncovered bases	86%	71%	77%	44%	80%
	Contigs uncovered	229,036	46,098	56,258	6,164	138,230
	% contigs uncovered	45%	16%	18%	4%	29%
	Contigs low-covered	479,893	211,863	255,174	119,398	430,590
	% contigs low-covered	93%	75%	80%	76%	89%
	Transrate score	0.13	0.26	0.25	0.21	0.42
	Optimal Transrate score	0.40	0.39	0.43	0.31	-
	Optimal improvement factor	3.13	1.51	1.67	1.45	-

Table 3.2: Transrate read metrics definitions

Metric	Explanation	Optimum
Reads	the number of read pairs provided	NA
Reads mapped	the total number of read pairs mapping	theoretically equal to fragments
% of reads mapped	the proportion of the provided read pairs that mapped successfully	theoretically 1.0 (see above)
Good mappings	the number of read pairs mapping in a way indicative of good assembly	equal to fragments
% good mappings	the proportion of read pairs mapping in a way indicative of a good assembly	1
Bad mappings	the number and proportion of reads pairs mapping in a way indicative of bad assembly	0
Potential bridges	the number of potential links between contigs that are supported by the reads	0
Bases uncovered by any reads	the number of bases that are not covered by any reads	0
% uncovered bases	the proportion of bases that are not covered by any reads	0
Contigs with uncovered bases	the number of contigs that contain at least one base with no read coverage	0
% contigs with uncovered bases	the proportion of contigs that contain at least one base with no read coverage	0
Contigs uncovered	the number of contigs that have a mean per-base read coverage of < 1	0
% contigs uncovered	the proportion of contigs that have a mean per-base read coverage of < 1	0
Contigs low-covered	the number of contigs that have a mean per-base read coverage of < 10	no specific optimum
% contigs low-covered	the proportion of contigs that have a mean per-base read coverage of < 10	no specific optimum
Optimal improvement factor	Optimal Transrate score/Transrate score	

3.3.1.4 Annotation

Of the 149,683 putative peptides in the peptide assembly, 89,233 returned a successful blast hit from the arthropod dataset of UniprotKB. Of these, 65,713 were subsequently mapped with at least one GO term.

3.3.1.5 The *Euphausia superba* Transcriptome Database

The EsT-DB can be accessed at www.krill.le.ac.uk. Access is currently restricted to authorized users; please contact bjh13@le.ac.uk to request access. Instructions for use are provided on the site.

3.3.2   Core circadian genes

3.3.2.1 Transcriptome mining summary

As a result of transcriptome mining using the fragments already sequenced, *Esperiod* was identified in full in a single contig, while *Estimeless* retrieved two fragmentary contigs that together allowed for the reamplification of the full coding sequence. For further analysis of these sequences, refer to Chapter Two.

No further isoforms of any of the core circadian genes were identified. Of the genes that had already been fully cloned only *Escry2* was present as a single complete contig, the others showing varying degrees of fragmentation.

3.3.2.2 Trinity vs a multi-assembler, multi-*k*-mer assembly

A single Trinity assembly produced 157,361 contigs, of which 20% were estimated by Transrate to be potentially coding contigs with an ORF. The total assembly contains 484,125 contigs with 28% estimated to be coding. Table 3.3 shows the results of mining each assembly for the complete, known coding sequences of *Euphausia*'s core circadian genes.

Table 3.3: Quality assessment of output from mining two assemblies using complete *E. superba* circadian gene coding sequences. *identification of the stop codon of *Estimeless* from the Trinity assembly required the user to notice a misassembly that truncated the protein sequence when viewed in the obvious reading frame. The query coverage range of 88 – 98% is accounted for by this misassembly; the latter figure is true only when using the corrected sequence.

Query	Trinity				Total Assembly			
	No. of contigs	Longest contig (bp)	Total query coverage	Covers start and stop codons?	No. of contigs	Longest contig (bp)	Total query coverage	Covers start and stop codons?
EsBMAL1	5	726	86%	Both	4	1150	88%	Both
EsCLOCK	5	1364	73%	Stop	3	2071	99%	Stop
EsCRY1	2	1100	98%	Both	2	1100	98%	Both
EsCRY2	2	1165	100%	Both	1	2207	100%	Both
EsPERIOD	4	1723	92%	No	1	4476	100%	Both
EsTIMELESS	5	1879	88-98%	Both*	2	3415	100%	Both

For all sequences queried, the total assembly produced coverage equal to or higher than that of Trinity, over an equal number of contigs or fewer. Coverage of *Esclock* was particularly improved in the total assembly, while *Estimeless* was subject to a misassembly by Trinity that could potentially hamper attempts to confirm the sequence through PCR reamplification. In the case of *Escry2* and *Esperiod*, the full coding sequence was identified in a single contig in the total assembly.

Further analysis suggested that querying the Trinity assembly using *Drosophila* CLOCK rather than the *Euphausia* ortholog – as would usually be the case for gene discovery purposes – would produce coverage of only 56% of the full *Esclock* coding sequence. The same query on the total assembly would still produce 99% coverage.

3.3.3 Regulatory genes

3.3.3.1 Transcriptome mining summary

Table 3.4 shows the primers used to confirm by PCR reamplification the correct sequences of certain regulatory proteins identified as orthologs from the tblastn output, or complete the coding sequence by RACE PCR. Table 3.5 shows the accession numbers for the proteins used to query the total assembly and the output contig details. Of the 17 regulatory sequences queried, only one failed to find an acceptable putative ortholog in the transcriptome, and only three of the successful hits were not putative full coding sequences.

Table 3.4: Primers used in reamplification of coding sequences of full putative circadian proteins or RACE extension of fragments. See Appendix III for thermal profile details. Profile 25 is a standard Q5 polymerase profile used for a number of reactions with only the annealing temperature changed, hence that detail is provided here.

Gene	Use	Forward	Reverse	Profile	Annealing temp (°C)	Extension Time (mm:ss)
EsckIIα	Full reamplification	GACCGGCCTACACATCTAATAC	GCTGCAGCGAATAGATGAGTTA	25	66	00:50
EsckIIβ	Full reamplification	ATGGCAGTTGCTCCGCGAT	TCAAAATGACACATCTCGCCAC	25	67	00:50
Escwo	Full reamplification	GAAGGTATAACATGCCCTTTA	CTCCCTATAGTTTTTTGACTTGATAAT	25	60	01:00
Esdbt	Full reamplification	TAATCCGCCCTTATTTGACTCT	CAATATTCTGCCTTGGGCATG	25	64	00:50
Esnemo	Fragment confirmation	TGTCCTATCAACGGTTCAGC	TGCACATTGGCTTATTGGAG	25	64	01:00
	RACE 5'		TGGTTCTTCTACACGCGCTAAACC	21		
	RACE 5' nested		AAGGAAGACCTTAACATGGTCGGC	10		
Espdp1	Full reamplification	TTCCCTCCTCAGGCCTTATT	CTATCCACTGGGCCAGGTAA	25	66	01:00
Espdh	Full reamplification	CCACAAGACCGTAACAACACA	ACGCATTTGCTACTTTCAGCC	25	66	00:50
Espdhr	Full reamplification	GTGTTTGGCCTGACATATGCT	GATACATGGGTGCTCTGGCA	25	66	00:50
Espp1	Full reamplification	GCTGCCATCTTGGACATTCT	CATTTAATAGCGGCCCTTCA	25	63	01:00
Esmts	Full reamplification	AATCTTGTCCAGGGGGAAAC	CTGCAACCGTTGAGCTGTAA	25	66	01:00
Eswbt	Full reamplification	CCGTGTGTTAAGGCTGTGAA	TGGAATCAGACGATTGTTGC	25	64	01:00
Estws	Full reamplification	GGGATGGCCGGTAAGTATCT	TATCGACCCACGCTACCAAT	25	67	01:00
Esshaggy	Full reamplification	CACCTGATCACTGCTGCATG	AAGTGTCGGCGGTAGTGTTG	25	67	00:50
Esvrille	Full reamplification	ACAGCAGCACATGTCCTC	CGAAGGACAAGTACAAGCAC	25	64	00:50

Table 3.5: Transcriptome mining: clock-related query protein details and *Euphausia superba* output contigs. Notes: [a] Query protein used is a *Mus musculus* ortholog. All others from *Drosophila melanogaster*. [b] EsNEJIRE, EsNEMO and EsSLIMB are 3' partial fragments.

Query Protein		*E. superba* transcript/protein identifications				
		Transcript			Protein	
Protein name	Accession	Contig ID	E-value	Length	Name	Length
AANAT	AAM68307	224038	9.00E-14	1312	EsNAT1	330
		25_comp13083_seq2	4.00E-07	1353	EsNAT2	216
CASEIN KINASE II α	AAN11415	29_comp3633_seq0	0	1321	EsCKIIα	350
CASEIN KINASE II β	AAF48093	25_comp1878_seq0	4E-119	1363	EsCKIIβ	219
CLOCKWORK ORANGE	AAF54527	k31.J5165333	2E-26	2267	EsCWO	707
CTRIP	AAF52092	31_comp9752_seq0	1E-167	8148	EsCTRIP	2152
DOUBLETIME	AAF57110	c67978_g1_i1	0	1699	EsDBT	345
JETLAG	AAF52178	21_comp16370_seq0	1E-15	1901	-	-
LARK	AAF50578	k31.J5152200	6E-59	1227	EsLARK	326
NEJIRE	AAF46516	31_comp5882_seq0	0	6924	EsNEJ	2072[b]
NEMO	AAF50497	k21.S7956922	0	2038	EsNEMO	456[b]
PAR DOMAIN PROTEIN 1 ε	AAF04509	k41.J3781389	8E-05	1835	EsPDP1	489
PIGMENT DISPERING FACTOR	AAF56593	k51.S2746663	8E-06	387	EsPDH1	74
		c68106_g1_i1	3E-05	531	EsPDH2	79
PIGMENT DISPERING FACTOR RECEPTOR	AAF45788	21_comp12084_seq0	3E-91	1588	EsPDHR	295
PROTEIN PHOSPHATASE 1	CAA39820	k51.J2636217	0	1870	EsPP1	329
PROTEIN PHOSPHATASE 2A - MICROTUBULE STAR	AAF52567	k61.J1884004	0	1927	EsMTS	309
PROTEIN PHOSPHATASE 2A - WIDERBORST	AAF56720	k51.S2657195	0	1769	EsWBT	461
PROTEIN PHOSPHATASE 2A - TWINS	AAF54498	21_comp997_seq2	0	3375	EsTWS	455
REV-ERBα[a]	NP_663409	k71.R1269094	1E-45	3326	EsE75	800
RORA	AAH03757	k31.R5154656	4E-48	2842	EsHR3	669
SHAGGY	AAN09084	27_comp3209_seq0	0	2222	EsSGG	415
SUPERNUMERARY LIMBS	AAF55853	25_comp34876_seq0	0	1843	EsSLIMB	612[b]
TAKEOUT	AAF56425	31_comp3978_seq0	1E-11	1086	EsTAKEOUT	247
VRILLE	AAF52237	25_comp4277_seq0	3E-43	2021	EsVRI	474

Table 3.6: blastp analyses of putative *Euphausia superba* clock-associated proteins against *Drosophila melanogaster* protein database (Flybase) and NCBI non-redundant protein database.

Query	Top Flybase hit				Top NCBI nr hit				
	Flybase No.	Associated gene name	E-value	% amino acid identity/similarity	Accession	Name	Species	E-value	% amino acid identity/similarity
EsNAT1	FBpp0291637	*CG13759-PD*	5E-23	19.5/31.3	XP_002071904	GK10244	*Drosophila willistoni*	4E-26	21.1/32.4
EsNAT2	FBpp0291637	*CG13759-PD*	5E-10	24.4/42.3	XP_001663122	AAEL012952-PA	*Aedes aegypti*	1E-16	26.8/51.1
EsCKIIα	FBpp0070041	*casein kinase IIa*	1E-162	81.9/86.4	EFN84867	Casein kinase II subunit alpha	*Harpegnathos saltator*	0	74.5/78
EsCKIIβ	FBpp0300427	*Casein kinase II β*	6E-114	85.5/90.5	ELR50200	Casein kinase II subunit beta	*Bos mutus*	5E-146	81.2/85.8
EsCTRIP	FBpp0310477	*circadian trip*	1E-173	33/43.7	ACC99349	ULF (TRIP12)	*Homo sapiens*	0	46/60
EsCWO	FBpp0081723	*clockwork orange*	8E-27	17.8/26.9	KDR16323	Hairy/enhancer-of-split with YRPW motif	*Zootermopsis nevadensis*	2E-61	23.3/31.9
EsDBT	FBpp0306615	*discs overgrown*	7E-157	58.1/64.8	AGV28719	Casein kinase 1 epsilon	*Eurydice pulchra*	0	94/94.8
EsE75	FBpp0074915	*Ecdysone-induced protein 75B*	2E-143	33.9/45.1	AGS94407	Ecdysteroid receptor E75	*Litopenaeus vannamei*	0	74.9/83.6
EsLARK	FBpp0076555	*lark*	2E-55	39.5/49.7	NP_523957	lark, isoform A	*Drosophila melanogaster*	2E-62	39.5/49.7
EsNEJ[a]	FBpp0305701	*nej*	0	-	KDR19833	CREB-binding protein	*Zootermopsis nevadensis*	0	-
EsNEMO[a]	FBpp0076474	*nemo*	0	-	NP_729316	Nemo, isoform E	*Drosophila melanogaster*	0	-
EsPDP1	FBpp0076495	*PAR-domain protein 1*	2E-35	27.5/38	EFN83234	Hepatic leukemia factor	*Harpegnathos saltator*	2E-39	24.6/30.8
EsPDH1	FBpp0084396	*Pigment-dispersing factor*	3E-05	30.7/36.8	JC4756	PDH related peptide precursor 79	*Penaeus sp.*	9E-27	67.5/76.2
EsPDHR	FBpp0309084	*Pigment-dispersing factor receptor*	9E-97	32.8/44.6	BAO01102	Neuropeptide GPCR B2	*Nilaparvata lugens*	1E-140	39/50.7
EsPP1	FBpp0306442	*Protein phosphatase 1a at 96A*	4E-178	90.6/94.5	EKC23784	PP1-alpha catalytic subunit	*Crassostrea gigas*	0	96.7/97.9
EsMTS	FBpp0310063	*microtubule star*	2E-176	92.9/97.4	KDR18186	PP2A catalytic subunit alpha	*Zootermopsis nevadensis*	0	95.8/98.4
EsWBT	FBpp0084575	*widerborst*	0	71.7/76	KPJ18201	PP2A regulatory subunit alpha	*Papilio xuthus*	0	83.6/90.1
EsTAKEOUT	FBpp0297106	*CG2016*	1E-32	30.7/54.1	ACO12182	Circadian clock-controlled protein precursor	*Lepeophtheirus salmonis*	1.00E-47	32.7/55.5
EsTWS	FBpp0081671	*twins*	0	74.7/83.6	AFK24473	PP2A regulatory subunit B	*Scylla paramamosain*	0	93.2/95.6
EsRORA	FBpp0297439	*Hormone receptor-like in 46*	3.45E-72	30/39.4	CAJ90622	HR3 isoform B1	*Blattella germanica*	0	48.1/59.8
EsSGG	FBpp0070450	*shaggy*	0	64/71.4	AEO44887	Shaggy	*Tribolium castaneum*	0	80.4/87.3
EsSLIMB[a]	FBpp0306059	*supernumerary limbs*	0	-	KDR19729	F-box/WD repeat-containing protein 1A	*Zootermopsis nevadensis*	0	-
EsVRI	FBpp0309715	*vrille*	4E-42	24.7/33.9	AAT86041	Vrille	*Danaus plexippus*	3E-68	37.3/48.4

3.3.3.2 Individual results

Casein Kinase II (CKII): A putative full protein for EsCKIIα-subunit, 350 amino acids in length, was identified. PCR reamplification revealed a 359 amino acid form that was taken as representative sequence. Domain analysis identified a serine/threonine protein kinase catalytic domain (Figure 3.4) while reverse-BLAST analysis showed the sequence was most similar to CKIIα in the ant *Harpegnathos saltator*. The same query against Flybase returned *D. melanogaster* CKIIα isoform A (Table 3.6).

A putative full protein for EsCKIIβ-subunit, 219 amino acids in length, was also identified. Domain analysis identified a casein kinase II regulatory subunit domain (Figure 3.5) while reverse-BLAST analysis showed the sequence was most similar to CKIIβ in the yak *Bos mutus*. The same query against Flybase returned *D. melanogaster* CKIIβ isoform I (Table 3.6).

```
MPLASRARVYADVNAHRPHEYWDYESHVIEWGQQDDYQLVRKLGRGKYSEVFEAVNINNNEKCVVKILKPVKKKKIKREIKIL
ENLRGGTNIITLQAVVKDPVSRTPALVFEHVNNTDFKQLYQTLNDYDIRYYLYELLKALDYCHSMGIMHRDVKPHNVMIDHEN
RKLRLIDWGLAEFYHPGQEYNVRVASRYFKGPELLVDYQMYDYSLDMWSLGCMLASMIFRKEPFFHGHDNYDQLVRIAKVLGT
EELYEYVEKYQIELDPRFNDILGRHSRKRWERFVHSENQHLVSPEALDFLDKLLRYDHQERLTAHEAMEHPYFYPIVKEQGRL
MSSPTPAPPGAPLSGVPE
```

Figure 3.4: Putative *Euphausia superba* CASEIN KINASE II α. Yellow highlighted region indicates a serine/threonine protein kinase catalytic domain.

```
MSSSEEVSWIAWFCGLRGNEFFCEVDEDYIQDKFNLTGLNEQVPHYRQALDMILDLEPDEEEEIPHQSDLIEQAA
EMLYGLIHARYILTNRGIAQMIEKYQAGDFGHCPRVYCENQPMLPIGLSDVPGEAMVKLYCPKCCDVYTPKSSRH
HHIDGAYFGTGFPHMLFMVHPEYRPKRPANQFVPRLYGFKIHPMAYQIQQQAAANFKAPMRVNYNNGKR
```

Figure 3.5: Putative *Euphausia superba* CASEIN KINASE II β. Green highlighted region indicates a casein kinase II regulatory subunit domain.

Clockwork orange (CWO): A putative full protein for EsCWO, 707 amino acids in length, was identified. Domain analysis identified a helix-loop-helix and Orange domain (Figure 3.6), while reverse-BLAST analysis showed the sequence was most similar to the CWO-like homolog Hairy/enhancer of split related with YRPW motif protein in the termite *Zootermopsis nevadensis*. The same query against Flybase returned *D. melanogaster* CWO isoform A (Table 3.6).

```
MLNMSTILSEDLHHPVPKSGYTTGGLGGLGMGLEGVSLGPGPPPSGGRIPAMFDHQAPRPHSPPLSSHPKDKRTHSGNKAQAN
ADRMPQLNPTSLHGHNRLTRRFNVEKKDPLSHRIIEKRRRDRMNNCLADLSRLIPAFYTKKGRGRIEKTEIIEMAIKYMKNLK
EHECQQLESCELATSQKEPPEGPSEPKQPPNQTDLYHMGYQDAINETMHFLNESEGLYPGDSLCVRLLNHLNKHQNDLMGGDV
SYSRAASGASSSSSGYHANGSSNSSSDGNGNASANSSDNKDTDITEESAYCSEDQPFPEDRDDKSSSRYDANQLREMLQRESF
QFLGGSDHSPEFTSGSYSQYLSSNNSSSGDDNGARLYKFKSNMKNRFTTDLEHSHQVRKRRRDSESSNGNYTDYEKDCGTPTG
LSSQPHNAPGKDVKRIYQNSPLLEKSLYSRKPIKEESISPNDVPIFALNKNGSFYIPLTIDHSLIAKGMARIIETTPVLHPV
SIFVNFSVPSSDESDKKFDNHYPSKQMPNISRDGEVDNRIPENSHDCPPDVCRDTQLSNPSYNDLKLRSGLQSLNVEDSENFD
KQNLFHPADIPHDTCHELPREHHKMIELRQGDHLPQGLRERIEPHFNGHQLPQEHTRDVRENRENLHHPDQYELQERHKLSQV
RVRRELKQSPHEFTEKTNYSGSSSYGRGPLPNWLQPNYQVKKL
```

Figure 3.6: Putative *Euphausia superba* CLOCKWORK ORANGE protein. Cyan highlighted region indicates a helix-loop-helix domain while pink indicates an Orange domain.

Circadian TRIP (CTRIP): A putative full protein for EsCTRIP, 2152 amino acids in length, was identified. Domain analysis identified a WWE domain common to TRIP12 peptides and predicted to be involved in ubiquitin-mediated proteolysis, and a HECTc domain, a feature of E3 ubiquitin-protein ligases (Figure 3.7). Reverse-BLAST analysis showed the sequence was most similar to the *Homo sapiens* ULF (TRIP12) protein while the same query against Flybase returned CTRIP (Table 3.6).

```
MAERGEGAAAGGSHSRATPSGTHTAKHRGRWSGGVAATAGGVEGLEAAGGSPLPIRKRKIGITEVLEPKRRTEETQIPGNFA
SGSNATSPSSSSSSTATANSNSNNSNNTNSRKRRRATSSGSSSELKEKDPNEVWAESPAKRQQRSSNIESQGQAAVSLAANAP
TTPTRNQESEKSSGGGILKSYPLRNRLRLSDGGGGGGGGGPGGGSPLGTTRHFEHHQSQSLPSSPSSQGVAFLADTFRRLRDS
RRSSQEQSSAQSTSKRNSQTKDPRQGPTGTWKRTHGTKKSPLSSKRSSRSKTNNTGSCASSSQQGGRGCGKVSFGSSSGGMSG
SHGPTQDGSSSSTGPAAAGHQQGAAPQPCTAPPPGSISSSVLATSGNAGNASLHTAPSTSTTSSLLPTASLATLGDSESEESD
MGRLEALLASRGLPPSLFGALGPRMQHILHRSVSSSISSKAQQLLVGLQASGDESEQLQAVIEMCQLLVMGNEDTLAGFPIKQ
VTPALIILLNMEHNFDMMNHACRALTYMMEALPRSSAVVVDAVPVFLEKLQVIQCMDVAEQSLTALDMLSRKHSKNILQAHGV
SACLMFIDFFSTGAQRAALSITANCCQNLTADEFHYVADSLQLLSNRLTSQGDKKCVESVCLAYSRLVDCFQNDPNRLQEIAG
HGLLKNIQQLLVMSPAVLSSVTFIMVMRMLSLMCNSCPQLAVELLKNNIANTLCYLLTNSTEPPPEQVELVSRTPQELYEITC
LIGELMPTLPSDGIFVIDGALQRPTGPPQETVTWQWQDDRAAWQNYSISDSRIIEMAHQSGEDEISLCIQGPAYTIDFHSMQQ
INDSSGTVRSVQRRLLPQGLQPPQQPDTQVLDARATCLQEDAALASNLIRSLFSLLYEVYSSSAGPAVRYKCLRALLRMIHFS
TPQLLQQVLKSQSVSSHLATMLSSHDLKIIVGAIQMTEILMQKLPEIFSIYFRREGVMHQIKRLCDPEITLGVVPSAKWSSPE
LSLIGGRSSSSSGGGTPISASPLLSTRPFNGNCLDASPSLPTPGPPLQSEDKLQSPQMRISDVLKRKRTTNRMRSGPFRKSRQ
EDSSSSSFSEFFMKNAGLGSGRESWRNMGGGRSKLMSGSNSGGGGGGGGGGSNAGTGVSPSSGKTASFLSSLNPSRWGRCNPAV
PTPPHDSPVSKEKLSNNASSVGSSGSSHSSSSHSSISGSSSSSNKEVVRVWVREQAQRLNEQYFTSELQGNTHPALSVLNRLI
AAIQQLDQEPLNALAALRDISAIVSNSDISPFEVIHSGLVRCLLAYLTGWDKSSSVSPSVSSTSDTSNSSPLQYSVSSQAGVE
TNGGTVLTNIATPNREVRLRNFLHVFLGCPVDSLSLGGVAWETIPVFSALVSKLNGCVNQLEQFPVKVHDLPGTALTGIGRGA
STSAIKFFNTHQLKCNLQRHPDCSRLRQWRGGPVKVDPLALVQAIERYLVIRGYGRLRDDDEENSDDENSDEDIDDTLAAVTV
NQTSSSHKLQFVIGDHVLPYNMTVYQAIRQFSSPSDASASEADTDTETPVSHSAIWLQTHTIYYRAVQEEEPTRSSGGRKGKA
GGSKPSPKKKADANLGDGSIVSGNHSVLEQYLQPNLPASVTVIDPSLEVLGLLRVLHAINRHYGTLYPPSVYPPPLPTTDFTN
LKLSAKATRQLQDPLVIMTGNVPPWLPQIAYACPFLFPFETRQLLFYAVSFDRDRALQRLMDSSPELNSVDSSERVTPRLERR
KRTVSRDDILKQAEAVMSDLASSRAMLEIQYENEVGTGLGPTLEFYALVSAELQKADLELWRGDTVTVEEGTTENPGKAVKYV
NVNHGLYPLPIARNIKSAHVTKLKSKFRFLGKFMAKAVMDSRMLDLGLHPCVYKWLLGEERSLGLGDVSSLDPSLGNTLNRLH
QTVMTKRRIQALSVTEDLTGAELQERLNQVTVDGCPVEDLALTFTLPGYPHIEFRKGGADMEVTINNIDQYLQLVIEWLVRDG
VWRQMEALREGFESVFPIAQLQVFYPEELEQLFCGSSDSSQWDVKQLAECCRPDHGYTHSSRAVKYLLQILAEYNGEQQRSLL
QFVTGSPRLPVGGFRSLSPALTIVRKTFEAGENPDEFLPSVMTCVNYLKLPDYSSIDIMRAKLEVAAREGQHSFHLS
```

Figure 3.7: Putative *Euphausia superba* CIRCADIAN TRIP protein. Blue/white text highlight indicates Armadillo repeat domain; red indicates WWE domain; teal/white text indicates HECTc domain.

Doubletime (DBT): A putative full protein for EsDBT was identified, 345 amino acids in length. Domain analysis identified a serine/threonine protein kinase catalytic domain

(Figure 3.8), while reverse-BLAST analysis showed the sequence was most similar to CKε in the sea louse *Eurydice pulchra*. The same query against Flybase returned *D. melanogaster* DCO (a synonym of DBT) isoform D (Table 3.6).

```
MELRVGNKYRLGRKIGSGSFGDIYLGTNISTGEEVAIKLECIKTKHPQLHIESKFYKMMAGGVGIPGIKWCGSEGDYNVMVME
LLGPSLEDLFNFCSRKFSLKTVLLLADQLITRIEYIHSKNFIHRDIKPDNFLMGLGKKGNLVYIIDFGLAKKYRDSRTHQHIP
YRENKNLTGTARYASVNTHLGIEQSRRDDLESLGYVLMYFNRGSLPWQGLKAATKRQKYERISEKKMQTPIEELCKGFPNEFA
TYLNFCRSLRFEEKPDYSYLRQLFRQLFHRQGFTYDYVFDWNMLKFGGTRNQENEVERRERGGSRPMGGGATSRIRHDTAVGG
VLPSPTAGKPLVI
```

Figure 3.8: Putative *Euphausia superba* DOUBLETIME protein. Yellow highlighted region indicates a serine/threonine protein kinase catalytic domain.

Jetlag (JET): While a putative full protein was identified as a candidate for EsJETLAG, reverse-BLAST analysis showed the sequence to be most similar to F-box/LRR-repeat protein 20 in various species, while searching Flybase showed the most to an F-box/LRR-repeat protein of unknown molecular function, CG9003. Bearing this in mind, the sequence was rejected.

Nejire (NEJ): A large 3' fragment was identified as a putative EsNEJ protein, 2072 amino acids in length. Domain analysis identified a bromodomain, a KAT11 histone acetylation domain, a zinc binding domain, a TAZ zinc finger, a SAM pointed domain and a WW (conserved with 2 Trp residues) domain (Figure 3.9). Reverse-BLAST analysis showed the sequence to be most similar to CREB-binding protein in *Zootermopsis nevadensis*, while Flybase returned NEJIRE (Table 3.6).

```
*QGPQQGLMKPGGLPGQNNPNQVVVFQSLPTDWLNLPNQPAQLPGGFGSVTAKPVKGTKEWQQAISSELRNHLVHKLVQAIFP
TPDPQAMLDKRMQNLVAYARKVEGDMFEQANSRPEYYHLLAEKIYKIQKELEEKRVQRKQMTNNRGDHPGPPGPPVPGVPLQN
QPPGPQGNQIPGLPPGPGTIGPRGPLNPRIPGLGNSMSGPGGPNMNQTAPQVSQHNIVGMPPGSVANTQFNNTIPSNSMTNST
PPFNNVGMPDGMPRPPPPYNNTPPEGSSVMNNQTNNCTSPALPVNSSVPTSVSIAGGQITPNNSMGGMNTTTNTVASAQSSVP
SSLTSTTPSTDGSLASSKPNMTTTSANSDSGTLVATSNHSINENNNGNSTPTNEIKQEEKPDIKNPFQNTIKMEETEGPEIKS
EIKSEPIEQDDVKDVKEMIKVEPKTEPMEEGSTTNTNSNSIGNDKTLVKEEPKSPIETPGGEDSNSMTGDQSCGKKIGKDGKQ
SGKQSQKGKGKQGNVAIRRPECTPKPPTPSPPPRPEPSMKKDRTPLFKPDELRHTLAPTMEKLYRQDPESIPFRQPVDPHALG
IPDYFDIIKKPMDLSTIKRKLDTGQYLDPWDYVDDVWLIFDNAWIYNRKTSRVYRYCTKLAEVFEMEIDPVMQSLGYCCGRKY
TFNPQVLCCYGKQLCTIPRDGKYFSYQNRYTYCAKCFNDIQGDTISLGDDPSQSALQIKKSQFQEMKNDALDLEPLVECGDCG
RKQHQICVLHLEQIWPHGFTCEGCLKKKAQKRKDNKFTAKKLPTTKLSNYLETRVNNFLKKKEAGAGEVYIRVVSSTDKAVEV
KGGMKTRFVDTGQLSPEFPYRAKALFAYEEIDGADVCFFGMHVQEYGSDAPNPNTRRVYLAYLDSVHFFKPRQYRTAVYHEIL
LGYLDYVKQLGYTMAHIWACPPSEGDDYIFHCHPNDQKIPKPKRLQDWYKKMLDKGIIERIVLDYKDIHKQAFEDNMKSPAEL
PYFEGDFWPNVMEESIKELDQEEEEKRKQEEAAAAEAAAQAALGGEEEEEVSCDGKKKGQMKKTNKKHNKKNTVQKSKAKKTT
GSQCNDLAQKIFATMEKHKDVFFIVRLHSAQSAVSLPPIQDPDPITVCDLMDGRDAFLTLASERHLEFSSLRRAKYSTMCMLY
ELHNQGQERFVYTCNSCKTHVETRYHCQECDDYDLCTTCFKTEGHPHKMIKLGFDLDDSAGSGDNKNLNPQEARRQSIQRCIQ
SLVHASQCRDANCRLPSCQKMKRVLSHTKSCKKKNNGGCAICKQLIALCCYHARNCKEQRCVVPYCYNIKQKLRQQEMQQRFQ
QQALMKRRMQQMNMRSIGGGPSISQPTPPQAVPQPMQSPIQQPQHPGVKPTQAPPPNVLQAVKMVQEEAARQTGPGTQVSVGK
GNFGGPVNSGQPQMMPPPNMPPQQVPRPQQMQPVNSMGVPMVQHGPMNSGPIGKPNSLMQQQQQQQQQQQQLQQQAKVPHQQQV
YYPDHPLLRERLEDDTFQINKSAVTFDPSGLSFLDSLVEVKKEVEEEVSVGSVCGSPAAPSPGVVSPAPGQQTSPQQTTPITT
TTSDILETVELLALEHARRDLQRVSDSLGIPRDPAQWSVDDVRSWLKLQASQLQLPPLMLDFWNLAGPALLKLSEAQFRQLAP
VGGELLHAQLEIWHQALRFSATPYYHHQLQQQAAAQQQQQQHQQQQQQQLHNMFMAHQHHLQQQQAAEALVAQQQQQQQQQQQQ
RANMASPVPSQQQQQQQQPPAQQQQQQQQPQQQVSGPLPGGWEQAVTPEGEVYFINHEDKSTSWFDPRIPRRLQTRNLVNGVQGT
QRAAPPAHSDTPQQQVSGLPPVAPGSGGGGLKAAQQQQQQQQQQTLRLQKLQLEHERLRMRQQEILKSEQSLRLRLREEATTN
AAAMQQQNMQNQENNNNNSNSNNNNAGDPFMSSGGPTATSNIDFHSRQESADSGLGLGCGSGDSGGLSIGSGSSYSMPHTPED
FLMAQDAPMNEGVEMKSVPGGASGGAATSMMHNDMGPATQELESMDSEDLVPTIQLNDDLSTDFLNDIIGSAKLDNMHTWL
```

Figure 3.9: Putative *Euphausia superba* NEJIRE protein fragment. Green/white text indicates bromodomain; violet/white text indicates KAT11 histone acetylation domain; dark yellow/white text indicates zinc binding domain; grey indicates zinc finger. Black/white text indicates SAM pointed domain; yellow indicates WW domain.

Nemo (NMO)*:* A large 3' fragment was identified as a putative EsNEMO protein, 456 amino acids in length, through the combination of two transcriptome fragments. RACE extension failed to produce any further sequence data. Domain analysis identified a serine/threonine protein kinase catalytic domain (Figure 3.10). Reverse-BLAST analysis showed the sequence was most similar to NEMO, isoform E, in *Drosophila melanogaster*. The same query against Flybase returned *D. melanogaster* NEMO isoform H (Table 3.6).

```
*VLSTVQPSQPYYAQVDPRGTPSQPSGLSGHGHHRSSSSSSQAAPTSQAPVGGTTLTQAPNSSPAQHPPPAPGAREAQEVTPD
RPIGYGAFGVVWAVTDPRNGERVALKKMPNVFQTLISSKRVFRELRMLCFFKHENVLSALDILQPPHIDFFQEIYVITELMQS
DLHKIIVSPQHLSADHVKVFLYQILRGLKYLHSARIIHRDIKPGNLLVNSNCVLKICDFGLARVEEPDESAHMTQEVVTQYYR
APELLMGARHYTQAVDVWSVGCIFGELLGRRILFQAQSPVQQLERITDLLGTPALEDMRSACEGAVSHMLRRAPKPPALPALY
TLSSHATHEAVHLLTQMLSFDPEKRLRVTEALAHPYLDEGRLRYHSCMCTCCASTNCGSRQYTHDFEPKAQQPFDDNWETDLR
SMSAVKERLHKFIMSHLARDRVPLCINPMSAAYKSFASSAST
```

Figure 3.10: Putative *Euphausia superba* NEMO protein fragment. Yellow highlighted region indicates a serine/threonine protein kinase catalytic domain.

Par domain protein 1ε (PDP1ε)*:* A putative full protein for EsPDP1 was identified, 489 amino acids in length. Domain analysis identified a basic region leucine zipper domain

(Figure 3.11), while reverse-BLAST analysis showed the sequence was most similar to hepatic leukemia factor in the ant *Harpegnathos saltator*. The same query against Flybase returned *D. melanogaster* PDP1 isoform D (Table 3.6). This was considered moderately supportive of this protein being part of the PDP1 protein family and EsPDP1 was tentatively accepted as such.

```
MDLVSALIKDRGEVSPTTTSPTTMHAFNSSHFGDLAASQEQQLQQKLRLQHQQQQHHQQQQQQRERLHPEDNSNNSNHQQQQQ
LHLNQQQQHQQQQQVRTHTHLQQIRQAIQLQQLQNQHSQHPYQQHQQPSPSQQQQQQHYATSSSPLQGTVANMPDSPNTPGSS
CAKEEDKTDEWQPYPESAFLGPTLWDKTLPYESQTALKQTTSSRQRPEVKLSEVHDSRGSTSIPSPSPPSSPNPRTFSDAMLE
YMDLDDFLDENNLHNGGNNNISNNNSNNSNISKNNNSSIGGHSVESSSTVPSDRSPQQSESPSSPDPTPTMMGTGLGACMGGA
MGAQLPPPQNMGLCSPPVMSPNDLMSPLSPGSPPDHSSMINGEYHSEHPHLESEHRGNFDPRTHQFTDDELKPHALVKKSRKQ
FVPCD<mark>LKDDRYWDRRFKNNAAAKRSRDARRCKENQIAMRANFLEKENNALTVEVEQANAIVDALKKRLSV</mark>YEAV
```

Figure 3.11: Putative *Euphausia superba* PDP1 protein. Green highlighted region indicates a basic region leucine zipper domain.

Protein phosphatase 1 (PP1): A putative full protein for EsPP1 was identified, 329 amino acids in length. Domain analysis identified a PP2A homologs catalytic domain (Figure 3.12), while reverse-BLAST analysis showed the sequence was most similar to PP1-alpha catalytic subunit in the Pacific oyster *Crassostrea gigas*. The same query against Flybase returned *D. melanogaster* PP1A isoform B (Table 3.6).

```
MADTDKLNIDSIIARLLEVRGSRPGKNVQ<mark>LTENEIRGLCLKSREIFLSQPILLELEAPLKICGDIHGQYYDLLRLFEYGGFPP
ESNYLFLGDYVDRGKQSLETMCLLLAYKIKYPENFFLLRGNHECASINRIYGFYDECKRRYNIKLWKTFTDCFNCLPVAAIVD
EKIFCCHGGLSPDLQSMEQIRRIMRPTDVPDQGLLCDLLWSDPDKDTMGWGENDRGVSFTFGAEVVAKFLHKHDFDLICRAHQ
VVEDGYEFFAKRQLVTLFSAPNYCGEFDNAGAMMSVDETLMCSFQILKPAD</mark>KKKFPYGGLNTGRPVTPPRGAANQKGKKK
```

Figure 3.12: Putative *Euphausia superba* PROTEIN PHOSPHATASE 1 protein. Cyan region indicates a PP2A homologs catalytic domain.

Protein phosphatase 2A (PP2A): *PP2A catalytic subunit MICROTUBULE STAR (MTS).* A putative full protein for EsMTS was identified, 309 amino acids in length. Domain analysis identified a PP2A homologs catalytic domain (Figure 3.13 A), while reverse-BLAST analysis showed the sequence was most similar to PP2A catalytic subunit alpha in the termite *Zootermopsis nevadensis*. The same query against Flybase returned *D. melanogaster* MTS isoform C (Table 3.6).

PP2A regulatory subunit — WIDERBORST (WDB). A putative full protein for EsWDB was identified, 461 amino acids in length, but PCR confirmation was unsuccessful. Domain

analysis identified a PP2A regulatory B subunit (B56 family) (Figure 3.13 B), while reverse-BLAST analysis showed the sequence was most similar to PP2A regulatory subunit epsilon in the louse *Pediculus humanus corporis*. The same query against Flybase returned *D. melanogaster* WDB isoform F (Table 3.6).

PP2A regulatory subunit — TWINS (TWS). A putative full protein for EsTWS was identified, 455 amino acids in length. Domain analysis identified a seven WD40 repeats (Figure 3.13 C), while reverse-BLAST analysis showed the sequence was most similar to PP2A regulatory subunit B in the crab *Scylla paramamosain*. The same query against Flybase returned *D. melanogaster* TWS isoform C (Table 3.6).

A

```
MEDKAQMKELDQWIEQLMECKQLAENQVKTLCEKAKEVLAKESNVQEVKSPVTVCGDVHGQFHDLMELFKIGGRSPDTNYLFM
GDYVDRGYYSVETVTLLVSLKVRYRERITILRGNHESRQITQVYGFYDECLRKYGNANVWKFFTDLFDYLPLTALVDSQIFCL
HGGLSPSIDTLDHIRALDRLQEVPHEGPMCDLLWSDPDDRGGWGISPRGAGYTFGQDISETFNHSNGLTLVSRAHQLVMEGYN
WCHDRNVVTIFSAPNYCYRCGNQAAIMELDDSLKYSFLQFDPAPRRGEPHVTRRTPDYFL
```

B

```
MSGSGGTFVDRIDPFAKRTLKKKPKRSQGSSRYRTTNDVELQPLPSLKDVPGSEQEDLFLRKLRQCCVGFDFLDPVADLKGKE
MKRSTLNELVDYITAGRGVLTEPVYPEIIRMIACNLFRTLPPSDNPDFDPEEDDPTLEASWPHLQLVYEFFLRFLESPDFQPA
IGKKVIDQKFVLQLLELFDSEDPRERDFLKTVLHRIYGKFLGLRAFIRKQINNIFLRFVYETEHFNGVGELLEILGSIINGFA
LPLKAEHKQFLIKVLIPLHKVKCLSLYHAQLAYCVVQFLEKDPTLTEPVIKGLLKFWPKTCSQKEVMFLGEIEEILDVIEPSQ
FVKIQEPLFKQIAKCVSSPHFQVAERALYFWNNEYIMSLIEENSNVILPIMFPALYRISKEHWNQTIVALVYNVLKTFMEMNS
KLFDELTASYKAERQREKKREREREELWKKLQALELNHQQGLQQQQ
```

C

```
MDHDDLMEECDLLAGNGEIQWCFSQVKGTLDDDVSEADIISCVEFSHDGDLLATGDKGGRVVIFQRDPSSKNCHPRRGEYNVY
STFQSHEPEFDYLKSLEIEEKINKIRWLKRRNPAHFLLSTNDKTIKLWKVSERDKRAEGYNLRDESGQIRDPSSLTSLRVPVL
KPMELMVEASPRRIFANAHTYHINSISINSDQETYLSADDLRINLWHMEVTDQSFNIVDIKPTNMEELTEVITAAEFHPLDCN
VFVYSSSKGTIRLCDQRVGALCDAHTKLFEEPEDPTNRSFFSEIISSISDVKFSNSGRYMISRDYLSVKVWDLHMETKPIETY
PVHEYLRSKLCSLYENDCIFDKFECCWSGDDSAIMTGSYNNFFRMFDRSGKRDITLEASRETAKPRTLLKPRKVCTAGKRKKD
EISVDCLDFNKKILHTAWHPSENIIAVAATNNLYIFQDKF
```

Figure 3.13: Putative *Euphausia superba* PROTEIN PHOSPHATASE 2 proteins (A) MICROTUBULE STAR, (B) WIDERBORST and (C) TWINS). Cyan region indicates a PP2A homologs catalytic domain, pink region a PP2A regulatory B subunit (B56 family) domain, and red indicates WD40 repeats.

Rev-erbα: (REV-ERBα): A putative full protein for EsE75 (following the naming convention of the *Rev-erbα* ortholog that was the top BLAST hit) was identified, 800 amino acids in length (Figure 3.14). Domain analysis identified a c4 zinc finger in nuclear hormone receptors domain and a ligand binding domain of hormone

receptors, while reverse-BLAST analysis showed the sequence was most similar to ecdysteroid receptor E75 in the whiteleg shrimp *Litopenaeus vannamei* and more specifically classified the c4 zinc finger domain as REV-ERB receptor-like. The same query against Flybase returned *D. melanogaster* Eip75B isoform C (Table 3.6).

```
MVQLDAPYSFGLKDMEEMIVAEFDGTTVLCRVCGDKASGFHYGVHSCEGCKGFFRRSIQQKIQYRPCTKNQQCSILRINRNRC
QYCRLKKCIAVGMSRDAVRFGRVPKREKAKILAAMQSVNARSQEKAVLAELEDDQRVTGAIIRAHIDTCDFTRDKVQSMLQQA
RHNPSYTQCPPTLACPLNPSPVPMQGQQQLLQDFSERFSPAIRGVVEFAKRLPGFQQLCQEDQVTLLKAGVFEVLLVRLAGMF
DSRTNTMLCLNGQLLRREALQSSPNARFLMDSMFDFSERVNHLNLTDAELAIFCAIVILAPDRPGLRNATLVENVQRRLVSCL
QAVTNKHHPENLTLHFDLLNKIPDLRTLNTLHSEKLLAFKMTEHTAAGTNWDDSRSSWSLGSEDKDSGVGSPVSSCHDDIMRS
PVSCSDSVYSGESGSSSESVCSNEVTGYGYTDLRPPFPNHRRRRDLSEGASSGDEASEASMKCPFTKRKSESPDDSGIESGTD
RSDKLSSPSVCSSPRSSIDEKSEEDDMPVLRRALQAPPIINTDLLMEEAYKPHKKFRAATLRREEEPHSSQSIGTSILAQTLS
QPPQMLNPSSLASTHSTLAASLASASISPSLAASHSTLARTLLEGPKMSENDLRRADMLHSMIMRTDARERLSSSSSRSSPGP
YYVPQACLERLQMPHSSQMPHSSWTCPPRSSSASSSTSDRSSPMQSSVTVQPRVHLLTTPTPSRYSSPMGMGAQPTCSPSVSS
SPQPQQQIQGLGAAPQRGSPSPMLLQVDIADSQPLNLSKKTPPPTPQEFVLEA
```

Figure 3.14: Putative *Euphausia superba* E75 protein. Blue/white text region indicates c4 zinc finger in nuclear hormone receptors domain, while dark green/white text indicates ligand binding domain of hormone receptors.

Rora (RORA): a putative full protein for EsHR3 (hormone receptor 3), 669 amino acids in length, was identified and assigned an alternative name on the basis of the blastp analysis. Domain analysis identified a DNA binding domain (c4 zinc finger in nuclear hormone receptors) for retinoid related orphan receptors (ROR) and a nuclear receptor ligand binding domain for ROR (Figure 3.15). Reverse-BLAST analysis showed the sequence was most similar to Hormone Receptor 3 isoform B in *Blattella germanica*, while Flybase returned Hormone receptor-like in 46 (Table 3.6).

```
MEAQIAQMDLLVELFGPGWDGDAGVDTTTPSSSTPSPSMHRPPSVEKKTNSIKGKAQIEIIPCKVCGDKSSGVHYGVITCEGC
KGFFRRSQSSVVNYQCPRQKNCVVDRVNRNRCQYCRLQKCLALGMSRDAVKFGRMSKKQREKVEDEVRYHKAQILAGRGPAGS
EPSPDTTHSLYEAQNPISSDIYTNPYYSEISPFTTTSFGGYNPQPTTNLSFTEFTENNVVDSTTFDSHINMRPSSMETMADSG
ALSPVVSSGIGSGVFGANRGLGEVGRSAGPGGGGVTHRGGVGGSICDVQATEVLIKQEPQTCFDMDTMQTTTTSSHLDDGSGF
SVASSISDKSLTAVDSTTYLPPTPRVTPMSPTSPPTLSQSHLHLQQQQHLLQQQQLHSQQQQQQQQQSGTVAGVTMLPDDEPC
APHPAQLSELLGRWVCDAHIRTCLYSSEQIEDCKRKQTLDLSKVTFYKNMAHEELWFDCAQKLTTVIQQIIEFAKAVPGFRKF
SQDDQIVLLKAGSFELAVLRMSRYYDVNQNCVVYGDTLLPMEAFLTTESAEMRLVNNVFEFSKTVAELKLTDTELGLYSALVL
LQPDRCGLKGVDEIAKLSEAVGRSLMQELEKTHTPIKGDVTVFAFLRARMPALRDLNQLHQEALSKFKRAMPDLQFSDLHKEI
FNVDS
```

Figure 3.15: Putative *Euphausia superba* HR3 protein. Grey/white text region indicates DNA-binding domain of Retinoid-related orphan receptors, while dark green/white text indicates ligand binding domain of hormone receptors.

Shaggy (SGG): A putative full protein for EsSGG was identified, 415 amino acids in length (Figure 3.16). Domain analysis identified a serine/threonine protein kinase domain, specifically a Glycogen Synthase Kinase 3 type, while reverse-BLAST analysis

showed the sequence was most similar to SHAGGY in the flour beetle *Tribolium castaneum*. The same query against Flybase returned *D. melanogaster* SHAGGY isoform A (Table 3.6).

```
MSGRPRTTSFAEGNKGPPSVTFPGMKISSKDGNKITTVIATPGQGSDRPQEVSYMDTKVIGNGSFGVVFQAKLCETGELVAIK
KVLQDKRFKNRELQIMRRLEHCNIVKLKYFFYSSGDKKEEVFLNLVLEFIPETVYKVARHHSKQKQTIPISYIKLYMYQLFRS
LAYIHSLGVCHRDIKPQNLLLDPETGVLKLCDFGSAKHLVRGEPNVSYICSRYYRAPELIFGATDYTTNIDVWSAGCVLAELL
LGQPIFPGDSGVDQLVEIIKVLGTPTREQIREMNPNYTEFKFPQIKSHPWQKVFRQRTPEDAITLVSRLLEYTPSARITPLQA
CAHKFFDELRNPDTRLPNNRELPPLFNFTEHEAKIQPELNSKLIPAHYRGEGGSSSSGGGGGGGGSSSVDTAEGAVAATVNDN
```

Figure 3.16: Putative *Euphausia superba* SHAGGY protein. Yellow highlighted region indicates a serine/threonine protein kinase catalytic domain.

Supernumerary limbs (SLIMB): A large 3' fragment from a putative EsSLIMB was identified, 612 amino acids in length (Figure 3.17). RACE extension failed to produce any further sequence data. Domain analysis identified a D domain of beta-TrCP, an F-box domain and 7 WD40 repeats, while reverse-BLAST analysis showed the sequence was most similar to F-box/WD repeat-containing protein 1A in the termite *Zootermopsis nevadensis*. The same query against Flybase returned *D. melanogaster* SLIMB isoform A (Table 3.6).

```
*RRSCSGCERVFSTVTAEMEADPILDDSLDSVESQGMGEVDGTGGGEVYSPMIPMTMIYDNNTSPMSVTTSGGGLLGNDVNSD
NDDNEPMTIIPITQITDVGGGGGRRKKESSAQFITEREVCINYFETWGEQDQLEFMEHLLARMCHYQHGHINAFLKPMLQRDF
ITLLPKKGLDHVAEKILSYLDAKSLREAELVCKEWHRVIADGLLWKKLIERKVRTDALWRGLSERKGWGTYLFKPRPGAQHPS
HTTYRKLYPKIIQDIQTIEANWRMGRHNLQRINCRSENSKGVYCLQYDDQKIVSGLRDNTIKIWDRSSLQCYKVSKVLTGHTG
SVLCLQYDERVIISGSSDSTVRVWDVTTGEMTNTLIHHCEAVLHLRFNNGMMVTCSKDRSIAVWDMVTPTEINLRRVLVGHRA
AVNVVDFDEKYIVSASGDRTIKVWSTGTCEFVRTLNGHKRGIACLQYRDRLVVSGSSDNTIRLWDIECGACLRILEGHEELVR
CIRFDNKRIVSGAYDGKIKVWDLQAALDPRAPAATLCLRTLVEHSGRVFRLQFDEFQIVSSSHDDTILIWDFLNCSPPDSSTD
AARSPTTADARYGDLSNAATDASLSPTIDDIS
```

Figure 3.17: Putative *Euphausia superba* SLIMB protein fragment. Red highlighted region indicates a D domain of beta-TrCP; teal/white text indicates an FBOX domain; green indicates ED40 repeats.

Vrille (VRI): A putative full protein for EsVRI was identified, 474 amino acids in length (Figure 3.18). Domain analysis identified a basic region leucin zipper domain, while reverse-BLAST analysis showed the sequence was most similar to VRILLE in the monarch butterfly *Danaus plexippus*. The same query against Flybase returned *D. melanogaster* VRILLE isoform E (Table 3.6).

```
MMVAETVKYPVNYPALIPLSPNLAGLAGGFPTCSSSMLHSSYPQSPHILDQKPLENPLHLQQLQLQQQQQQQQEDSLSSEQGP
IGLGANGLVLPNSLDLAAIRKKEMFSQRKQREFIPDAKKDDSYWDRRRRNNEAAKRSREKRRFNDMVLEQRVIELAKENHIYM
AQLSAIKDKYGIDGDSMINVEQIMQTFPNTEQILACTKRSKYANIGNGLLNPSSPGSPSTSASSPAEHRDSLDNHNMDMDLSP
NHFYSNSHQSSSHHRGESPYELSRHQYPPTSSSSNFYEQSALNLSARPASPSPPVTHMEYSSSDEMSRRSPSDEQQGSCLPLK
LRHKTHLGDKDVAASLLALHYIKTEPRDSHADPADDSDDRDSGLGYSSSSSSSSSSADYIRPENTSPSMMGNSLSLSQRVQHAH
DQDIQEITDYNRENTSSNLLNGNNTHLKTELERLSSEVATLKYLLVSRPRPESDSDGSR
```

Figure 3.18: Putative *Euphausia superba* VRILLE protein. Turquiose highlight indicates basic leucine zipper domain.

3.3.4 Clock controlled genes

3.3.4.1 Transcriptome mining summary

Table 3.5 shows the accession numbers for the proteins used to query the total assembly and the contig details of those identified as orthologs from the tblastn output.

3.3.4.2 Individual results

Lark (LARK): A putative full protein for EsLARK was identified, 326 amino acids in length. Domain analysis identified two RNA recognition motifs and a zinc finger domain (Figure 3.19). Reverse-BLAST analysis showed the sequence was most similar to LARK isoform A *Drosophila melanogaster*. The same query against Flybase returned *lark* (Table 3.6).

```
MPVRGNTFKIFIGNLSDRTTGQDLRQLFEEYGTVVEADAVGGKNFGFVHMEKEEERKTAVESLNGHTLHDREIVVEASTGNRK
GGNKRTKIFVGNLHKDTTSGEIRDLFEAHGTVNEADVLSNFAFVHMESEDDARKAIESLDGTELHGLRLRVQESTSRVRQNAG
MGDRDSCYRCGSRGHWSKDCPRDGGRRGDRFEGGRGGGRFDSYGGPPPRGYDRDRMMRGMRDDPYDRFSRYGDDPYARRPLPP
RPLPPMRDDPYERRPLPPRLRDDPYDRRPMMPPMRDDPYERRQPMGMDYMRYGRRSRSPPPRMERSYGRFPPPPPPF
```

Figure 3.19: Putative *Euphausia superba* LARK protein. Olive green indicates RNA recognition motif; black/white text indicates a zinc finger domain.

Pigment dispersing hormone (PDH): Two putative full proteins for EsPDH were identified, 79 and 74 amino acids in length and likely to be isoforms from the same gene. Domain analysis identified a 5' signal peptide domain and a 3' pigment dispersing hormone domain in both (Figure 3.20 A), while reverse-BLAST analysis showed the sequences were most similar to pigment dispersing hormone related peptide precursor 79 in shrimp genus *Penaeus*. The same query against Flybase returned *D. melanogaster* PDF isoform H (Table 3.6). The shorter of the two sequences is identical to that designated as PDH-β by Toullec *et al*. (2013). While the search using

the *D. melanogaster* PDF sequence produced only EsPDH-β, a subsequent search using the designated *Euphausia* PDH-Lα sequence identified by the same team showed it to be present in the transcriptome. This appears to be a more complete and likely more correct form than that published, appearing to be a putative full protein and showing a high degree of sequence similarity to the same protein in the related species *Euphausia crystallorophias* (Figure 3.20 B). **Receptor *(PDHR)*:** A putative full protein for EsPDHR was identified, 295 amino acids in length. Subsequent cloning and sequencing revealed a likely assembly or sequencing error had produced a truncated protein and the full putative EsPDHR was 471 amino acids in length. Domain analysis identified a 5' signal peptide domain, a hormone receptor domain and a 7 transmembrane receptor (secretin family) domain (Figure 3.21), while reverse-BLAST analysis showed the sequence was most similar to GPCR BR in the brown planthopper *Nilaparvata lugens,* a member of the secretin-like GPCRs along with *Drosophila* PDF receptor. The same query against Flybase returned *D. melanogaster* PDFR isoform D (Table 3.6).

A

MRSTMSIILVLVAMSVALTYAQSDLKYSDREVVAHLAGQILRYVHGGWADGAQVPHKRNSELINSLLGLPKVMNDAGRR

MRSTMSIILVLVAMSVALTYAQSDSQYSDREVVAHLAGQILRFVHGAQVPQKRNSELINSLLGLPKVMNDAGRR

B

```
EsPDH-Lα_Toullec    ---ELVWSSSVCSLPSLPSMTWLTQSLTEQDLSVVASLASKILDVVAGSENGVDSDMQAL
EsPDH-Lα_Est-DB     MRASVVFLGLF--LAFIAIHDVITQSLTEQDLSVVASLASKILDVVAGSENGVDSDMQAL
EcPDH-Lα            MRASVVFLGLF--LAFIAINDVVAQSLTEQDSIVVASLAAQILDVVAGSENDGDSGVQAV
                     .:*: . .  *  :     ::*******  *****::********** ** :**:

EsPDH-Lα_Toullec    KRNSGTINSMLGLPRTYNLRRMMMNAGRR
EsPDH-Lα_Est-DB     KRNSGTINSMLGLPRTYNLRRMMMNAGRR
EcPDH-Lα            KRNSGTINSMLGLPRTYNLRRMMMNAGRK
                    ****************************:
```

Figure 3.20: A) Putative *Euphausia superba* PDH isoforms 1 (top) and 2 (bottom). Red highlight indicates signal peptide region; yellow indicates pigment dispersing hormone domain. B) Multiple alignment of *E. superba* PDH-Lα as determined by Toullec *et al*. (2013) and as extracted from the transcriptome described here (Est-DB), with PDH-Lα in *E. crystallorophias*. Red indicates signal peptide region.

```
MIHLYTSVCLYALLIMVDGNETSDLGFGIANQEHCVAYHANTTIQGPGWCSAAWDQIVCWPPSPPAYTAMLPCPPLKGVDPEQ
WAYRTCGPGGQWLGRPPEYLPPGWTNYTLCFIPEIRDLMAKLYKTSQEDAKRKLRVAESSRILETVGLSISLAAILISLIIFS
HFRVLRNNRTRIHWHLFVAIKIQLVVRLTLYIDQAITRAKLNIGFNHVGIDNTPVLCEGFYVLLEYARTAMFLWMFIEGMYLN
NMVTLSVFHDKPNYRFYYLLGWGLPVLMTASWAVVTAYKYNHTVCWWGYSFTSFFWILEGPRLAIILVNLIFLLNILRVLIMK
LQESLSSESQQIKDRVDIMKKAVRAAIVLLPLLGITNSIQMIHSPLDRDLISFAAWTYSTTMLTSFQGFFVALIYCFMNNEVR
SALRKSLANYHSRRNIDEHHRNSLASYNLYRSAAQLSYSSKHITASRASRDSSSQC
```

Figure 3.21: Putative *Euphausia superba* PDHR protein. Red highlight indicates signal peptide region; purple/white text highlight indicates a hormone receptor domain; green indicates a 7 transmembrane receptor (secretin family) domain.

Takeout (TO): A putative full protein for EsTO was identified, 247 amino acids in length. Domain analysis identified a juvenile hormone binding protein domain (Figure 3.22). Reverse-BLAST analysis showed the sequence was most similar to circadian clock-controlled protein precursor in the salmon louse *Lepeophtheirus salmonis*. The same query against Flybase returned *CG2016* (Table 3.6).

```
MKVLVAILVVGIASHVKSQFASSLRQCRVDNNNQLNSCLVQTMESLRPHLRTGVPELSLPVLEPMFIPNLNFRQGNGAVNINA
VFRNVEIRGLSSFNTTYIDADPRSQTLNVGLYIPELTVTGNYELDGLLILLPIEGRGPFWTTFTGINANGVGNIDIAGQGPTE
RLMVSSVYVNFDINNMRVKLDNLFNGDPILGEAVHLFLNENGKEVLAEIKPEIQRRLNELVQKVMNDAFSQLPVDTFIRRN
```

Figure 3.22: Putative *Euphausia superba* TAKEOUT protein. Yellow highlight indicates juvenile hormone binding protein domain.

3.3.5 Hormones, neuropeptides and receptors

3.3.5.1 Transcriptome mining summary

Table 3.7 shows the accession numbers for the neuropeptide/receptor proteins used to query the total assembly and the contig details of those identified as orthologs from the tblastn output. The following detail focuses on those genes that are of particular interest in circadian research.

3.3.5.2 Individual results

Insect-like arylalkylamine N-transferase (AANAT): Two putative full proteins were identified, 330 (EsNAT1) and 216 (EsNAT2) amino acids in length (Table 3.6). Domain analysis identified an N-Acyltransferase superfamily domain in both (Figure 3.23). Reverse-BLAST analysis of EsNAT1 showed the sequence was most similar to GK10244 in *Drosophila willistoni* while the same query against Flybase returned *CG13759-PD* (Table 3.6) – both are N-acyltransferases in the GNAT superfamily. The same process for EsNAT2 returned AAEL012952-PA in *Aedes aegypti* and *CG13759-PD* respectively.

The blastp analysis of EsNAT1 suggested a possible misassembly, with no matches at the N-terminus except for low identity hits with bacterial peptidases. Considering this, the phylogenetic tree was generated using EsNAT1 edited to remove the N-terminus. The phylogenetic tree generated (Figure 3.24) suggests that while these two contigs are part of the GNAT superfamily, and the *Euphausia* contigs cluster with those specifically identified in *Daphnia pulex* as AANATs (Schwarzenberger and Wacker, 2015), their status as typical insect-like arylalkylamine N-transferase is uncertain, hence the naming convention adopted.

```
MWSGEGNPLTAAASKVPGVSVNQEKASKSFFGMVDSVASCLRKYSVPIAGKKEEEESAEVAPSNAAVQQGKEQGANQPAEVVS
QPIEEPEMPELPVEPELPPDVIYMHDGQIAFKILTEEYIDKAVVLLCNQFFKDEPLGKALKLESPREVDHWLSKVLPHMIAHG
VSLMAIDESPEGEGRLVGVAINNVKLQGAVGGPDDFLNWIDPQKDPKMFRIISFLSHLAQDIDFYGNYGVDKFFNFEMLNVDK
AYGGRGLASMLVEQSLKLAQRLNFRLCLVETTGIFSAKIFSRHGFKTIREVSYGNFRENGKLLFPNTGIHTSARVCIKLLE


MEYTLLSEKDWDDVKPLLKDSFLVREPTMLALHLKPEEVMNLYSVIVDMTLNSGISYGARDTGTGKLVGFMLNKISLLKDQEM
EKNFEWDGTPGENSFIEVFRDLFKEIDLFAQNGVERIMDFCCLTVDPSYKGKGIARKLVELSEQKGIEEGCPLGKVEASNTIT
QHIWKQLGYQIHKMLDFQEYNKEKGKEVFDVKAAAPTTGWHCMSKRLDGK
```

Figure 3.23: Putative *Euphausia superba* NAT proteins EsNAT1 (top) and EsNAT2 (bottom). Pink highlight indicates an N-Acyltransferase superfamily domain.
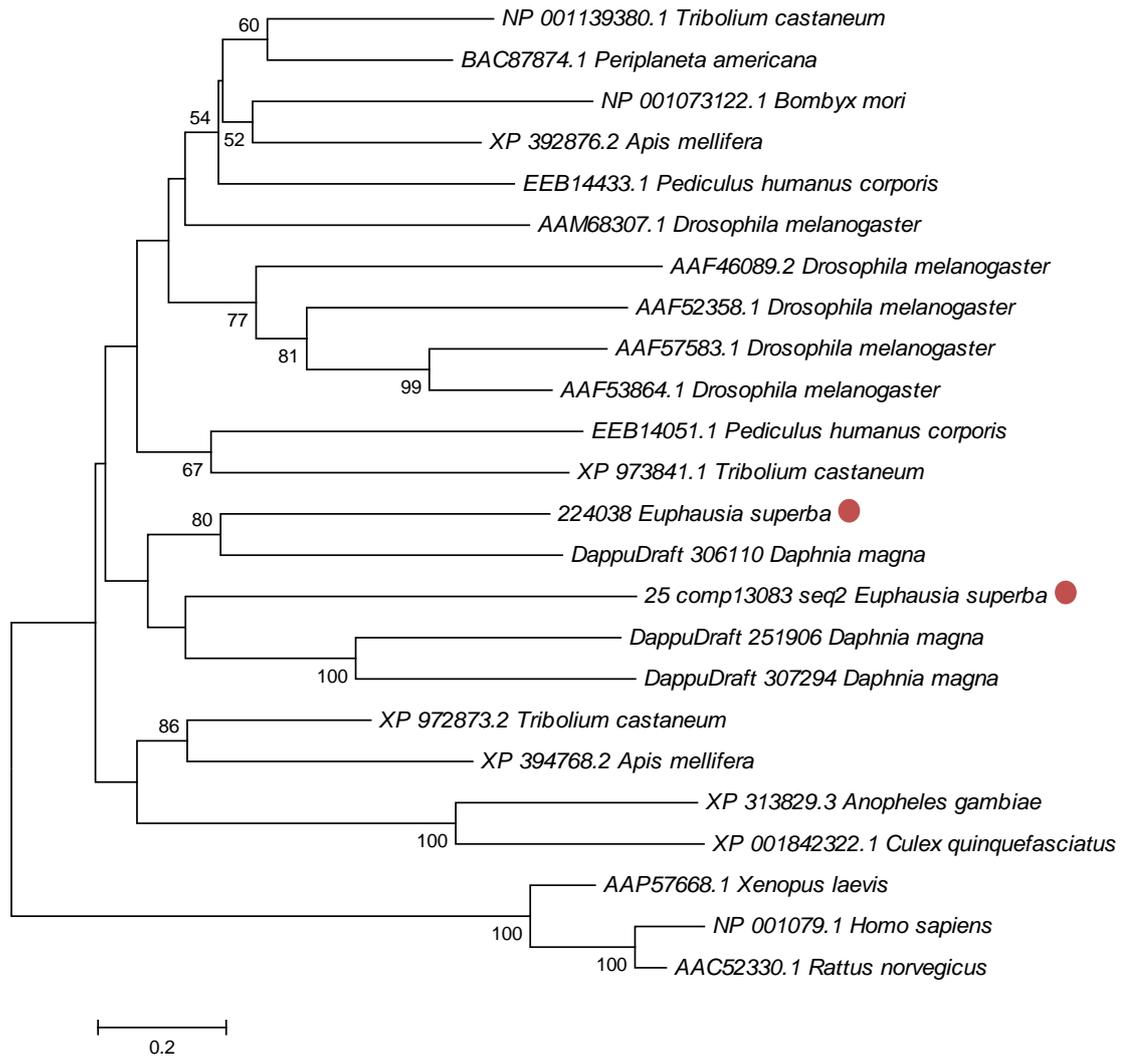
Figure 3.24: Phylogenetic tree generated from insect-like AANAT protein sequences and putative *Euphausia superba* candidates (red circles).

Red pigment concentrating hormone (RPCH): No candidates were identified for this neuropeptide using the *Daphnia pulex* ortholog (accession BJ935589; Table 3.7 A). However using the RPCH protein sequence for *Euphausia crystallorophias* reported by Toullec *et al*. (2013) a putative full protein sequence for the preprohormone was identified, 104 amino acids in length. Reverse-BLAST analysis showed that the sequence was most similar to RPCH in the river prawn *Macrobrachium rosenbergii*, with an E-value of $1e^{-16}$.

Crusctacean hyperglycaemic hormone (CHH). A putative full protein was identified for the preprohormone EsCHH, 131 amino acids in length (Table 3.7 A). Reverse-BLAST analysis revealed EsCHH to show most similarity to CHH in the crab *Charybdis japonica*.

3.3.5.3 Other neuropeptides and receptors

The full results from mining for putative *Euphausia superba* neuropeptides and receptors are shown in Table 3.7. Of 31 preprohormones queried 20 returned results, all of which were accepted as putative orthologs of the query sequence. The receptor search returned only five hits that could be accepted as orthologs without deeper sequence analysis. These five were allatostatin A and C receptors, crustacean cardioactive peptide receptor, FMRFamide-like peptide (other), and pyrokinin receptor, this latter with a top BLAST hit of neuromedin-U receptor 2 in *Tribolium castaneum*, an alternative name for pyrokinin 1 receptor C.

Table 3.7: A) preprohormone query proteins and B) receptor query proteins with *Euphausia superba* output contigs and subsequent blastp analysis against NCBI non-redundant protein database.

A

Peptide family (subfamily)	EsT-DB blastp of query sequence					NCBI NR Database blastp of contig		
	Accession	Contig	Size (nt)	E-value	Type	Accession	Description	E-value
Allatostatin A	ABS29318	27_comp22909_seq0	1239	4.00E-19	Fragment	BAF64528	allatostatin precursor protein *Panulirus interruptus*	1E-60
Allatostatin B	AAF49354	27_comp14162_seq0	710	2.00E-09	Fragment	AFV91538	B-type preproallatostatin I *Pandalopsis japonica*	3E-15
Allatostatin C	AAF53063	23_comp57987_seq0	757	4.00E-05	Fragment	BAO00935	Allatostatin-cc *Nilaparvata lugens*	5E-13
Allatotropin	AAB08757	-	-	-	-	-	-	-
Bursicon alpha	EFX87546	21_comp33157_seq0	552	2.00E-60	Fragment	AKJ74864	bursicon alpha subunit *Penaeus monodon*	4E-72
beta	EFX87749	21_comp28222_seq0	721	7.00E-37	Fragment	AKJ74865	bursicon beta subunit *Penaeus monodon*	9E-63
CHHamide	NP_001097784	25_comp21895_seq1	829	6.00E-04	Fragment	EFX80320	CCHamide-like precursor *Daphnia pulex*	3E-08
Corazonin	AAB32283	29_comp12038_seq1	365	2.00E-04	Fragment	Q9GSA4	Corazonin precursor-related peptide *Galleria mellonella*	2E-06
Crustacean cardioactive peptide	EFX70015	27_comp31900_seq0	685	2.00E-02	Complete	ABB46291	crustacean cardioactive peptide *Carcinus maenas*	4E-33
Crustacean hyperglycemic hormone	ABQ41269	k41.R3771434	691	5.00E-24	Fragment	ACN87216	crustacean hyperglycemic hormone precursor *Charybdis japonica*	8E-27
Calcitonin-like diuretic hormone	ACX46386	c66980_g1_i1	608	1.00E-47	Fragment	ACX46386	prepro-calcitonin-like diuretic hormone *Homarus americanus*	1E-53
Corticotropin-releasing factor-like diuretic hormone	AAF54421	-	-	-	-	-	-	-
Ecdysis-triggering hormone	AAF47275	-	-	-	-	-	-	-
Eclosion hormone	AAA29310	21_comp1319_seq1	536	3.00E-10	Fragment	BAO00951	eclosion hormone 2 *Nilaparvata lugens*	2E-16
FMRFamide-like peptide (myosuppressin)	ACX46385	265178	343	2.00E-34	Fragment	BAG68789	myosuppressin-like neuropeptide precursor *Procambarus clarkii*	9E-36
FMRFamide-like peptide (neuropeptide F)	AEC12204	23_comp4135_seq0	700	3.00E-26	Complete	AEC12204	preproneuropeptide F I *Litopenaeus vannamei*	5E-27
FMRFamide-like peptide (short neuropeptide F)	AAU87571	k61.S126078	378	4.00E-08	Fragment	ETN63818	short neuropeptide F precursor *Anopheles darlingi*	2E-13
FMRFamide-like peptide (sulfakinin)	ABQ95346	-	-	-	-	-	-	-
FMRFamide-like peptide (other)	BAE06262	63534	521	6.00E-06	Fragment	AAR19420	FMRFamide-like peptide precursor *Periplaneta americana*	1E-08
Insulin-like peptide	AAS65048	-	-	-	-	-	-	-
Inotocin	ABX52000	c73008_g1_i1	538	3.00E-17	Fragment	BAO00906	arginine vasotocin precursor *Paralichthys olivaceus*	2E-24
Leucokinin	AAF49731	-	-	-	-	-	-	-
Neuroparsin	ACO11224	31_comp27103_seq0	522	1.00E-15	Fragment		Neuroparsin-A precursor *Caligus rogercresseyi*	4E-16
Orcokinin	ACB41787	k51.J2644345	765	2.00E-59	Fragment		Orcokinin-like peptide 4 Precursor *Procambarus clarkii*	1E-77
Periviscerokinin/pyrokinin	NP_001104182	-	-	-	-	-	-	-
Proctolin	CAD30643	-	-	-	-	-	-	-
Red pigment concentrating hormone	BJ935589	-	-	-	-	-	-	-
Adipokinetic hormone	EFX68649	-	-	-	-	-	-	-
RYamide	EDP28140	-	-	-	-	-	-	-
SIFamide	BAC55939	23_comp6805_seq0	935	7.00E-25	Fragment	Q867W1	SIFamide precursor *Procambarus clarkii*	8E-24
Tachykinin-related peptide	ACB41786	c73859_g1_i1	1289	6.00E-37	Complete	BAD06363	preprotachykinin B *Panulirus interruptus*	1E-37

B

Peptide family (subfamily)	EsT-DB blastp of query sequence					NCBI NR Database blastp of contig		
	Accession	Contig	Size (nt)	E-value	Type	Accession	Description	E-value
Allatostatin A	AAF45884	c46254_g1_i1	467	6.00E-44	Fragment	AAK52473	allatostatin receptor *Periplaneta americana*	3E-51
Allatostatin B	EFX87704.1	29_comp4995_seq4	1754	9.00E-105	Complete	XP_002061030	GK10663 *Drosophila willistoni*	1E-121
Allatostatin C	AAF49259	21_comp79404_seq0	601	4.00E-56	Fragment	AIY69138	allatostatin receptor 3 *Neocaridina denticulata*	4E-100
Allatotropin	ADX66344.1	27_comp4908_seq5	2026	8.00E-24	Complete	BAO01065	neuropeptide GPCR A15 *Nilaparvata lugens*	9E-120
Bursicon	ABA40401	25_comp16561_seq	5328	0.00E+00	Complete	KOC60881	Lutropin-choriogonadotropic receptor *Habropoda laboriosa*	0E+00
CHHamide	NP_611241	21_comp152_seq4	1717	1.00E-107	Fragment	KFM78411	[Phe13]-bombesin receptor *Stegodyphus mimosarum*	2E-121
Corazonin	AAF49928	29_comp42008_seq0	1334	5.00E-22	Fragment	AIT57587	crustacean cardioactive peptide receptor *Cancer borealis*	8E-146
Crustacean cardioactive peptide	AAO66429	c56113_g1_i2	1504	2.00E-107	Fragment	AIT57587	crustacean cardioactive peptide receptor *Cancer borealis*	4E-146
Crustacean hyperglycemic hormone	*No known receptor*	-	-	-	-	-		-
Calcitonin-like diuretic hormone	AAN16138	21_comp12084_seq0	1588	2.00E-23	Complete	BAO01102	neuropeptide GPCR B2 *Nilaparvata lugens*	3E-89
Corticotropin-releasing factor-like diuretic hormone	AAF58250	21_comp45712_seq0	942	4.00E-34	Fragment	XP_002063277	GK21476 *Drosophila willistoni*	7E-38
Ecdysis-triggering hormone	ABN79653	25_comp39064_seq0	1555	1.00E-66	Fragment	XP_002001513	GI21940 *Drosophila mojavensis*	1E-79
Eclosion hormone	*No known receptor*	-	-	-	-	-		-
FMRFamide-like peptide (myosuppressin)	NP_647713	21_comp92052_seq0	342	5.00E-18	Fragment	XP_001848191	g-protein coupled receptor *Culex quinquefasciatus*	6E-24
FMRFamide-like peptide (neuropeptide F)	AAK50050	23_comp82370_seq0	476	1.00E-41	Fragment	KFM78023	Neuropeptide Y receptor type 6 *Stegodyphus mimosarum*	3E-43
FMRFamide-like peptide (short neuropeptide F)	AAF49074	21_comp67157_seq0	617	3.00E-50	Fragment	EFN70952	Prolactin-releasing peptide receptor *Camponotus floridanus*	5E-71
FMRFamide-like peptide (sulfakinin)	EDS26978	29_comp86831_seq0	404	5.00E-27	Fragment	XP_001866739	CCK-like GPCR *Culex quinquefasciatus*	2E-32
FMRFamide-like peptide (other)	AAF47700	21_comp76838_seq0	355	7.00E-28	Fragment	NP_001280540	FMRFamide receptor *Tribolium castaneum*	2E-30
Insulin-like peptide	EFX62637	27_comp10497_seq1	2676	1.00E-60	Complete	KOC63013	Tyrosine kinase receptor Cad96Ca *Habropoda laboriosa*	2E-154
Inotocin	NP_001078830	c56113_g1_i2	1504	4.00E-36	Fragment	AIT57587	crustacean cardioactive peptide receptor *Cancer borealis*	4E-146
Leucokinin	AAF50775	29_comp86831_seq0	404	6.00E-50	Fragment	ADM47603	kinin receptor *Ixodes scapularis*	3E-54
Neuroparsin	*No known receptor*	-	-	-	-	-		-
Orcokinin	*No known receptor*	-	-	-	-	-		-
Periviscerokinin/pyrokinin	AAS65092	25_comp17524_seq0	1394	1.00E-54	Fragment	NP_001290187	neuromedin-U receptor 2-like *Tribolium castaneum*	5E-105
Proctolin	NP_572183	21_comp53150_seq0	447	2.00E-35	Fragment	KDR19658	FMRFamide receptor *Zootermopsis nevadensis*	1E-52
RPCH/adipokinetic hormone	ACD75498	29_comp42008_seq0	1334	3.00E-29	Fragment	AIT57587	crustacean cardioactive peptide receptor *Cancer borealis*	8E-146
RYamide	*No known receptor*	-	-	-	-	-		-
SIFamide	ABB96223	21_comp152_seq4	1717	8.00E-26	Fragment	KFM78411	[Phe13]-bombesin receptor *Stegodyphus mimosarum*	2E-121
Tachykinin-related peptide	AAA28722	21_comp152_seq4	1717	2.00E-33	Fragment	KFM78411	[Phe13]-bombesin receptor *Stegodyphus mimosarum*	2E-121

3.3.6 Abundance

The relative abundances of the contigs of interest in the transcriptome are shown in Table 3.8. The 25 *k*-mer Bridger assembly contained 117,543 contigs with a TPM ranging from 0 to 5,638. 12,064 contigs had a TPM of 10 or greater, while 5,462 had a TPM of 20 or greater. Of the core circadian genes, *Escry2* clearly showed the highest transcript abundance. The full *Esbmal1* sequence returned a value slightly outside of the range calculated for its fragments, while the others fell within their respective ranges.

The result of extrapolating from the linear regression model to estimate the number of expressed genes well-supported in the transcriptome is shown in Figure 3.25. Based on the Y-intercept an estimated 6,372 genes are expressed.
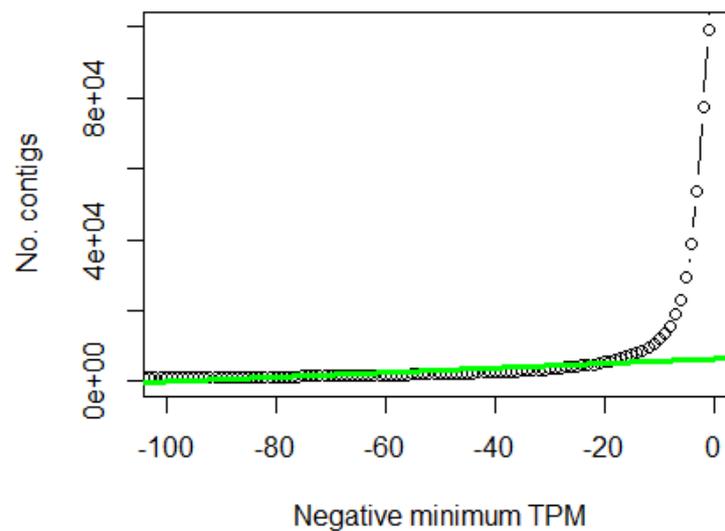
Figure 3.25: Number of contigs with a negative minimum TPM between 0 and -100 (all contigs in the transcriptome present at 0). Fitted line represents linear regression of this data filtered to exclude contigs with negative minimum TPM > 10.

Table 3.8: Abundance values derived from RSEM alignment of reads to the 25 *k*-mer Bridger transcriptome. To calculate abundance for the bolded sequences these were manually added to the transcriptome, the associated fragments (listed above each bolded line) removed and the calculation rerun.

Gene type	Contig ID	Circadian gene	Length	Effective length	Expected count	TPM	FPKM
Core	comp69792_seq0	*Esbmal1*	539	362.92	10	1.84	1.81
	comp85796_seq0	*Esbmal1*	285	109.55	4	2.43	2.4
	comp53867_seq0	*Esbmal1*	334	158.32	6	2.53	2.49
	comp62946_seq0	*Esbmal1*	266	90.64	5	3.68	3.62
	comp42461_seq0	*Esbmal1*	397	221.06	15	4.52	4.45
	Full sequence	***Esbmal1***	**2157**	**1980.89**	**37**	**1.25**	**1.23**
	comp32596_seq1	*Esclock*	660	483.89	5.6	0.77	0.76
	comp32596_seq0	*Esclock*	657	480.89	24.4	3.38	3.33
	comp26173_seq0	*Esclock*	501	324.95	22	4.51	4.44
	comp31442_seq0	*Esclock*	378	202.13	16	5.28	5.19
	comp15945_seq0	*Esclock*	1364	1187.89	136	7.63	7.51
	Full sequence	***Esclock***	**4044**	**3867.89**	**277**	**4.78**	**4.7**
	comp63811_seq0	*Escryptochrome1*	538	361.92	10	1.84	1.81
	comp21135_seq0	*Escryptochrome1*	1100	923.89	56	4.04	3.98
	comp34358_seq0	*Escryptochrome1*	305	129.46	9	4.64	4.56
	Full sequence	***Escryptochrome1***	**1644**	**1467.89**	**43**	**1.95**	**1.92**
	comp5516_seq0	*Escryptochrome2*	2226	2049.89	656	21.34	20.99
	comp20192_seq0	*Esperiod*	4213	4036.89	241	3.98	3.92
	comp28718_seq0	*Estimeless*	1332	1155.89	56	3.23	3.18
	comp16463_seq0	*Estimeless*	1554	1377.89	105	5.08	5
	comp17408_seq0	*Estimeless*	1884	1707.89	161	6.29	6.18
	Full sequence	***Estimeless***	**4222**	**4045.89**	**282**	**4.65**	**4.57**
	comp37177_seq0	*Estimeout*	878	701.89	27	2.56	2.52
	comp19904_seq0	*Estimeout*	3064	2887.89	198.28	4.58	4.5
Regulatory	comp2399_seq0	*Estwins*	3381	3204.89	2381.5	49.54	48.74
	comp4277_seq0	*Esvrille*	2021	1844.89	989.4	35.76	35.18
	comp5250_seq3	*Eswiderborst*	6035	5858.89	856.58	9.75	9.59
	comp9456_seq1	*Escircadian trip*	4576	4399.89	680.67	10.31	10.15
	comp3462_seq2	*EsckIIA*	4107	3930.89	1207.11	20.47	20.14
	comp1878_seq0	*EsckIIB*	1363	1186.89	1899	106.68	104.96
	comp16115_seq1	*Esclockwork orange*	2301	2124.89	64.83	2.03	2
	comp4349_seq1	*Esdoubletime*	2480	2303.89	181.38	5.25	5.16
	comp2884_seq0	*Ese75*	3602	3425.89	2514.64	48.94	48.15
	comp21231_seq0	*Eslark*	1196	1019.89	71	4.64	4.57
	comp1421_seq0	*Esmicrotubulestar*	2263	2086.89	2728.2	87.16	85.76
	comp3122_seq0	*Esnejire*	5204	5027.89	731	9.69	9.54
	comp21158_seq0	*Esnemo*	1995	1818.89	135	4.95	4.87
	comp1987_seq0	*Espdh*	709	532.89	518.32	64.85	63.8
	comp12996_seq0	*Espdhr*	1588	1411.89	155.15	7.33	7.21
	comp39942_seq0	*Espdp1*	1079	902.89	35	2.58	2.54
	comp1492_seq1	*Espp1A*	3028	2851.89	3609.06	84.38	83.01
	comp7384_seq0	*Esrora*	3293	3116.89	170.67	3.65	3.59
	comp32873_seq0	*Esrpch*	533	356.93	32	5.98	5.88
	comp1237_seq1	*Esshaggy*	7343	7166.89	5346.18	49.74	48.93
	comp34876_seq0	*Essupernumerary limbs*	1843	1666.89	70	2.8	2.75
	comp3527_seq0	*Estakeout*	1090	913.89	847	61.79	60.8

3.4 Discussion

3.4.1 A suite of clock genes

3.4.1.1 Overview

Detailed above is an analysis of a comprehensive suite of putative clock-related genes and their protein products in *Euphausia superba*. Full coding sequences were obtained for 17 proteins that may contribute to generation and maintenance of the Antarctic krill's molecular clock, of which 11 of particular interest were confirmed by PCR, while full coding sequences were obtained for five proteins that may be involved in the output of the clock, two confirmed by PCR. Beyond this, coding sequences for 21 neuropeptide preprohormones including the clock-relevant red pigment dispersing hormone (RPCH) and crustacean cardioactive peptide (CCAP) were identified.

3.4.1.2 Crustacean clockwork

Six years ago, as noted by Strauss and Dircksen (2010), only a single crustacean clock gene had been sequenced; the c*lock* ortholog of the giant river prawn *Macrobrachium rosenbergii*, *Mar*-Clk (Yang *et al.*, 2006). Many more have since followed: the core circadian genes of the isopod *Eurydice pulchra* (Zhang *et al.*, 2013): a cryptochrome of *Euphausia superba* (Mazzotta *et al.*, 2010) and the amphipod *Talitrus saltator* (O'Grady and Wilcockson, unpublished; accession AFV96168); and putative complete circadian systems in *Daphnia pulex*, the copepods *Tigriopus californicus* and *Calanus finmarchicus* (Nesbit and Christie, 2014; Christie *et al.*, 2013a), and, described first here and in Chapter Two, *Euphausia superba*. RNA-seq has been described as a "revolutionary tool" (Wang *et al.*, 2009) and this promise is fulfilled in the speed at which it has enabled the discovery of large sets of coding sequences in a fraction of the time.

The discovery of so many clock and clock-related genes leaves the researcher with an abundance of possibilities. While the intractable nature of *Euphausia* as a model organism prevents complete reenactment of the work that resulted in the detailed documentation of *Drosophila's* circadian system, many molecular studies performed to identify the functions of these genes in *Drosophila* can be repeated with the krill

genes, such as the promotion or repression of expression or activity and the identification of phosphorylation targets.

3.4.1.3 More in hope than expectation

Not all of the regulatory genes detailed in this chapter are expected to contribute to *Euphausia*'s biological clock. While keeping an open mind about the possibility of the involvement of mammalian Rev-erbα and Rora orthologs, it seems more likely that the krill clock will tend towards the generalities of the invertebrate system.

That said, recent research has identified a role in the *Drosophila* circadian clock for the closest ortholog of mammalian Rev-erbα, the ecdysone-induced protein 75 (Kumar *et al.*, 2014), or E75. This nuclear receptor is expressed in clock cells, and shows transcript cycling. Overexpression or RNAi knockdown wipes out the locomotor rhythm, and the former dampens *per* and *Clk* oscillations and protein abundance. Finally, E75 appears to repress the transcription of *Clk* through interactions with its promoter. It is findings such as these that underpin the logic of such a wide-ranging search of the *Euphausia* transcriptome; new components, or a role for genes previously thought not to apply, can be discovered at any time.

3.4.2 Clock controlled genes

3.4.2.1 The obvious candidates...

A number of candidates for *Euphausia* clock controlled genes have been identified through transcriptome mining (Table 3.6). While it is not necessarily expected that *Eslark* and *Estakeout* will be found to be as tightly bound to the biological clock as their *Drosophila* orthologs they are certainly a reasonable place to start, and perhaps even better are certain peptides such as EsRPCH, EsCHH and EsPDH and their receptors. Each of these latter has been linked to the output of the clock in other crustacean species (Sosa *et al.*, 1994; Kallen *et al.*, 1990; Verde *et al.*, 2007), and other neuropeptides identified here have been implicated in insect circadian systems, such as crustacean cardioactive peptide (CCAP) (Sehadová *et al.*, 2007).

3.4.2.2 …and the rest

Beyond that, many other neuropeptides have been identified and as with the others it remains to be seen if they are under the control of the clock. The search for associated receptors was much less successful, though it does not necessarily mean that there are not valid contigs to be obtained from the transcriptome. Further work is necessary to investigate each in greater detail than the broad sweep of transcriptome mining allows, and it may be that the query sequences used may need to be tailored more specifically towards crustacean data.

3.4.3 The database

3.4.3.1 A high quality assembly

Efforts to produce a high quality *de novo* assembly were rewarded with an excellent success rate in searching for genes of interest. This was reflected in the less fragmented nature of the core clock gene contigs (Table 3.3), which produced full sequences coding for EsPERIOD and EsCRY2. This also meant that a reasonable search query (*Drosophila* CLOCK) was capable of retrieving 99% of the coding sequence for EsCLOCK in the total assembly compared to 56% in a single *k*-mer Trinity assembly. This result is due to less conserved regions having been successfully incorporated into longer contigs containing more conserved areas that were then identified by the blastp query; sequence data that could never be identified using homology is retrieved when the contigs are more complete.

Further evidence of the assembly quality comes in the form of the rating assigned to the total assembly by Transrate. Smith-Unna et al. (2015) conduct a meta-analysis of published *de novo* assemblies and conclude that an assembly with a Transrate score of 0.22 (with an optimised score of 0.35) would be superior to 50% of the assemblies deposited in the NCBI Transcriptome Shotgun Assembly Database - the total assembly Transrate score is 0.42 (Table 3.1 A). It is interesting to note that the Trans-ABySS and SOAPdenovo-Trans assemblies alone also fulfill the criteria needed to surpass 50% of the assessed assemblies, and it is this author's experience from early usage of each that the former in particular is an excellent assembler when employed in a multi-*k*-mer approach. Bridger (Chang *et al.*, 2015) also contributed many of the best candidate

contigs to the final output, and while it appears to produce a high number of poor quality contigs as well (Table 3.1 B) this could be limited to very low k-mer assemblies: a trial run using a *k*-mer of 25 (as used by Trinity) was rated by Transrate as superior to all others in terms of initial score and second only to the Trans-ABySS multi-*k*-mer assembly for optimised score. Bridger employs the same logic as the scaffolding option of SOAPdenovo-Trans in utilizing paired read information to further extend assembled contigs and join shorter ones together, a process that may simultaneously explain the superior N50 of its outputs and its apparent tendency to produce more misassembled and/or low-coverage reads than the other assemblers. Use of Bridger followed by Transrate's identification of quality contigs could prove to be an optimum approach for producing a single *k*-mer assembly for downstream analyses (see 3.4.3.2).

Another small point of comparison is between the PDH-Lα sequence identified here against that published by Toullec *et al.* (2013) (Figure 3.20 B). It is clear from the alignment with that of *Euphausia crystallorophias* that the Est-DB assembly has generated a corrected, complete transcript for that gene.

Finally, numerous assembly attempts were made before settling on the pipeline described above. In early assemblies not only were the mined sequences more likely to be fragmented (*Esperiod*) or incomplete (*Estimeout*, *Escwo*) but obvious misassemblies were sometimes identified, such as palindromic contigs that contained partial coding sequences that read the same in both directions. These were not apparent however in the total assembly, as each assembler used and each *k*-mer run increased the chances of the assembly of the correct contig, while downstream stripping of redundancy and quality assessment of contigs meant only the best candidates remained.

3.4.3.2 Abundance

The total assembly and its derivations, the coding and peptide assemblies, were designed to maximise gene discovery potential. To this end, the assembly pipeline aimed for a high number of unique contigs at the expense of other considerations. While a number of processes were employed to reduce the number of redundant or erroneous contigs in order to maintain assembly quality, all were run at 100% identity so as to never remove a truly unique contig. As a result the assemblies still

undoubtedly contain many duplicate contigs with minor differences due to assembler error, or fragments with overlap but without containment, resulting in the removal of neither. This is unavoidable and was considered an acceptable compromise in the pursuit of quality coding sequences.

Going forward, one may wish to map reads to the transcriptome to assess differential expression between samples or identify cycling transcripts. It is likely that the transcriptome would benefit from further processing prior to this in order to further remove redundancy. Using CD-HIT-EST (Fu *et al.*, 2012; Li and Godzik, 2006) with a sequence identity threshold of less than 1 (100%) would more effectively remove contigs with minor differences at the cost of the loss of some possibly correct sequences, but with a likely benefit to downstream analyses thanks to the removal of ambiguity for the mapping software.

Alternatively, as conducted here, a single assembly can be used as an alternative for downstream analyses. After identifying the coding sequences in the total assembly, one can use these to mine a single assembly for the same sequences, albeit sometimes fragmented or with slight variations. As shown in Table 3.8 a fragment of a transcript can reasonably serve as a proxy for the whole, as the abundance metrics are adjusted for transcript length and fragment values are comparable to those obtained for the complete sequence.

The Est-DB is the output of sequencing data derived from a single sample, and thus the read alignment is not especially informative, allowing us simply to estimate the relative abundance of each transcript. It is clear that with 20.99 TPM *Escry2* is the most highly expressed of the core circadian genes, while the others are of a similar level, ranging from 1.25 – 4.78 TPM. This degree of abundance also appears to be the range at which fragmented contigs are likely to appear, with *Esnemo* (4.95 TPM) and *Esslimb* (2.75 TPM) failing to assemble completely despite the attempts made to assemble as complete a transcriptome as possible.

The estimate of expressed genes (6,372) produced by the linear regression (Figure 3.25) may be lower than the reality, but perhaps not by much. The logic in calculating this is that an influx of poorly supported transcripts appear at low TPM and therefore

by excluding them and going by the trend prior to their appearance one can estimate the number of well-supported genes *including* those lowly expressed. However the trend is distinctly flat in this transcriptome which may lead to an underestimate. Taking into account the number of transcripts with a reasonably stringent cut-off of TPM > 10 is 12,064, the true number may lie between this figure and the linear regression estimate.

3.4.4   Conclusions

A high quality *de novo* transcriptome was assembled using a multi-assembler, multi-*k*-mer pipeline. The majority of sequences sought were clearly identified as putative full transcripts and few queries retrieved nothing of value. A putative circadian system has been identified and described in *Euphausia superba*, as well as a broad collection of genes that could be characterized as the output of the clock.

With the Est-DB, available at www.krill.le.ac.uk a transcriptome resource has been created that will be of use to researchers from many disciplines; not only those directly studying *Euphausia superba* but those with an interest in molecular genetics, ecology and evolutionary biology, not necessarily restricted to Crustacea. It has already been employed in the identification of an ancient conserved non-coding element in the 5' region of the *Not1* gene (Davies *et al.*, 2015) and the high success rate when searching for clock-related sequences bodes well for its wider usefulness.

Chapter 4   Is the locomotor activity of the amphipod *Parhyale hawaiensis* under endogenous control?

4.1   Introduction

4.1.1   Crustaceans in chronobiology

4.1.1.1 Not-so-model organisms

As documented in Chapter One, many crustacean species have been studied by chronobiologists, encompassing a number of major classes. They have particular characteristics that make them attractive for certain research assays – an obvious and reliable behavioural pattern like the swimming behaviour of *Eurydice pulchra*, or a physiological response such as the ERG amplitude of *Procambarus clarkii*. But few approach the broad flexibility and utility of the classic model organisms such as *Caenorhabditis elegans*, *Mus musculus*, *Danio rerio* or *Drosophila melanogaster*. Some are relatively large, making space and/or time an issue, and this no doubt can have an impact in achieving sufficient experimental power in study design. They can be difficult to culture, meaning animals must be sampled in the wild or collected and transported for experimentation, and often breed only during a particular time of year, imposing a seasonal limitation if one wishes to control for age or study a particular cohort. Crucially for modern research methods, genetic manipulation tools are sorely lacking.

Studies of *Nephrops norvegicus,* for example, are typically conducted on low numbers of wild caught animals (Aguzzi and Chiesa, 2005; Aguzzi *et al.*, 2009; Sbragaglia *et al.*, 2015). In *Euphausia superba*, the seasonal regression to a juvenile state (Kawaguchi *et al.*, 2007) makes age determination incredibly difficult. *Eurydice pulchra* ticks many boxes – a small isopod that can be kept in the lab, possesses clear circadian and circa-tidal rhythms and in which RNAi has been successfully implemented (Zhang *et al.*, 2013) – but it has a seasonal reproductive period spanning April – August and long maturation times, with juvenile females maturing from July and becoming ovigerous in April of the following year (Jones, 1970).

4.1.1.2 A malacostracan *melanogaster?*

Amphipods are an order of crustaceans from the Malacostraca class that also encompasses krill, isopods and decapods. They have been dubbed "the flies of the sea" (Browne *et al.*, 2005) for their near-ubiquity in aquatic environments, species diversity and broadly scavenging nature. Such an analogy, while made only in reference to ecological traits, is prone to draw the attention of those familiar with the success story of the fruit fly *Drosophila melanogaster* as a model organism. If amphipods are the flies, does a *Drosophila* of the sea lie within the clade? In the past two decades, a crustacean model organism has been established and developed to an extent that suggests the answer may be a positive one.

4.1.2 *Parhyale hawaiensis*, the organism

4.1.2.1 Distribution

The amphipod *Parhyale hawaiensis* (Dana, 1853; hereafter *Parhyale* unless otherwise specified) is found in shallow water intertidal zones. It is considered a pantropical organism with specimens identified in Brazil, Haiti, Florida, Bermuda, the Belgian Congo, California, Ecuador, Hawaii, India, Australia and more (Shoemaker, 1956), although it has been suggested that these recordings may represent a species complex (Myers, 1985).

4.1.2.2 Ecology, diet and morphology

Much of what is known about aspects of *Parhyale's* life history traits such as lifespan, growth and reproduction come from modern lab studies after the establishment of the species as a model organism. Early literature has a distinctly broad ecological focus on identification and habitat in works encompassing rudimentary details of many dozens of species in a particular ecosystem or region (Barnard, 1965; Rao, 1972).

Further insight can be taken from a study on feeding behaviour conducted on specimens caught in Northern Australia (Poovachiranon *et al.*, 1986). As is typical for *Parhyale,* judging by entries in the above cited ecology literature, the area from which they were taken was dominated by mangroves, mainly *Rhizophora* species. Being detritivores favouring decaying mangrove leaves the animals were found amongst the leaf litter of these trees, and were by far the most abundant amphipod sampled (>

75%) living in dense populations of up to almost 7,000 individuals per m^2. The waters in which it was found were subject to wide fluctuations in salinity from freshwater sources, from 27 to 41 ‰, and the amphipod proved tolerant of an even wider range of 5 – 40 ‰ in the lab.

Parhyale shows clear sexual dimorphism, mature males (10 - 12 mm) distinctly bigger than females (7 – 8.5 mm) and possessed of a much enlarged second gnathopod (Figure 4.1; Shoemaker (1956)) that allows easy identification and separation of adults. The second gnathopod is used by the male to initiate and maintain an amplexus – a pre-mating pairing – up to the point of egg fertilisation.
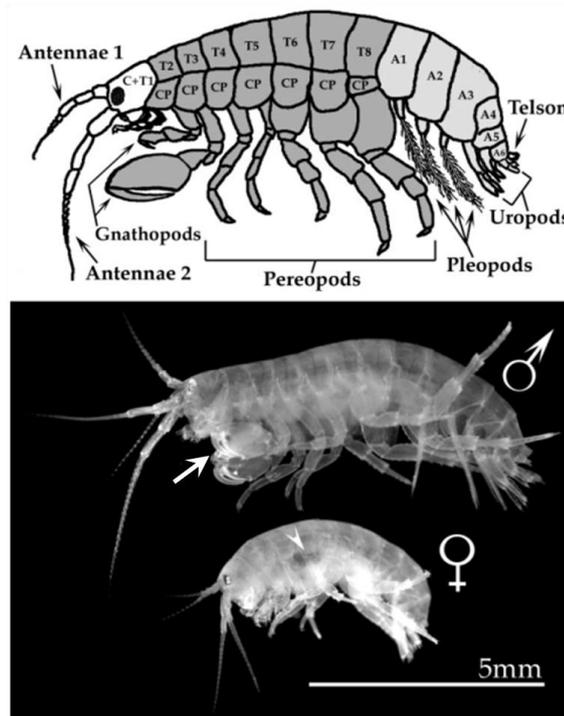
Figure 4.1: Top - schematic of male adult *Parhyale hawaiensis*. C – cephalon; T(n) – thoracic segment; CP – coxal plate; A(n) – abdominal segment. Bottom – male and female morphology. Males possess an enlarged second gnathopod (arrow), females a ventral brood patch (arrowhead). Modified from Browne *et al.* (2005).

4.1.3 *Parhyale hawaiensis*, the model organism

4.1.3.1 Establishment

In 1997 the founding isolate from which all subsequent lab cultures have been derived (Rehm *et al.*, 2009c) was taken from the marine filtration system of Chicago's John G.

Shedd Aquarium by members of the Patel lab, then of the University of Chicago. The story as told by a current member of the lab, now located at the University of California, is that the group were searching for a new crustacean to work with and visited the aquarium with the specific aim of finding one with the low maintenance characteristics of the best model organisms – hardy and unfussy in dietary requirements – and their focus was therefore directed towards the waste system rather than the display tanks (Jarvis, 2014). The amphipod found thriving on the filters was only identified as *Parhyale hawaiensis* after establishment (Gerberding *et al.*, 2002).

4.1.3.2 Culture and reproduction

In the lab *Parhyale* is typically maintained in artificial seawater or filtered natural seawater with a layer of coral or other calcium carbonate gravel provided as a substrate and to buffer pH. Aeration is important in large tanks, but smaller cultures can do without provided the water is shallow – one or two cm above the substrate (Ramos, personal communication). At least partial coverage is recommended to minimise salinity increase through evaporation. Cultures should be maintained at 20 – 25°C; the higher temperature is optimal for growth.

Parhyale's main food source in nature consists of decaying plant matter. While a recommended menu of kelp granules, fatty acids, plankton and vitamins has been put together, a much more basic regime of carrot is sufficient to maintain small populations (Rehm *et al.*, 2009c). It is not an obligate detritivore, however, and cannibalism will become overt if food is withheld for long.

The egg-to-egg generation time is typically two months, and reproduction occurs year round, every two to three weeks once a female reaches maturity, with an average brood size of six from a range of one to 30 in larger animals (Browne *et al.*, 2005). *Parhyale*'s mating system enables confidence in parentage and allows the generation of inbred lines: during reproduction the male clasps and guards the female until her next moult, at which point he deposits his sperm and releases her; the female then releases her eggs into a ventral brood pouch (Figure 4.1) before the cuticle hardens, fertilising them in the process. Fertilised eggs, which are large enough to allow

microinjections, can subsequently be removed from the brood pouch at the one-cell stage and raised to hatching over 10 days in sterilised seawater. A direct developing organism, *Parhyale* hatchlings are morphologically like tiny adults. Once raised to maturity, they can be crossed with siblings or parents.

4.1.3.3 Experimental toolkit

Parhyale has been described as "the most powerful available crustacean model for developmental genetic and molecular cell biology studies" (Kontarakis and Pavlopoulos, 2014). Examples of protocols that have been established include *in situ* hybridisation of digoxigenin and fluorescein labelled RNA probes (Rehm *et al.*, 2009b), antibody staining to determine protein localisation (Rehm *et al.*, 2009a), cell ablation (Chaw and Patel, 2012), RNAi knockdown (Liubicich *et al.*, 2009) and conditional gene misexpression through the use of inducible heat-shock elements (Pavlopoulos *et al.*, 2009). The latter study, in which misexpression of the Hox gene *Ultrabithorax* transformed the anterior maxilliped appendages into thoracic leg-like structures, employed the use of the *Minos* transposable element to randomly generate transgenic animals, a technique previously established in *Parhyale* (Pavlopoulos and Averof, 2005). This approach has been expanded upon to develop a system dubbed integrase-mediated trap conversion (iTRAC), in which *Minos*-generated gene traps can be converted through the integration of secondary constructs. In principle, this technique enables the development of binary expression systems such as the GAL4/UAS system employed to such powerful effect in *Drosophila melanogaster*. Further suggested applications include genetic cell ablation, expression markers and gene knockout (Kontarakis *et al.*, 2011).

Very recently, targeted mutagenesis has been established in *Parhyale* with the successful implementation of the CRISPR/Cas9 system, injection of early stage embryos generating somatic loss-of-function mutations in Hox genes (Martin *et al.*, 2015). The system has also been employed to insert a GFP reporter into the *Antp* Hox gene that was able to reveal expression patterns at late developmental stages where the cuticle typically prevents *in situ* or immunohistochemistry analyses (Serano *et al.*, 2016).

4.1.4 Aims and objectives

Parhyale hawaiensis meets many of the criteria that define a modern model organism, but that does not automatically qualify it for adoption in a particular field of research. *Drosophila melanogaster* may possess the toolkit to break apart the molecular clock, but without behavioural assays such as eclosion and locomotor activity one cannot easily observe the effects of such tinkering – there must be a measurable phenotype of some kind. The laboratories that have adopted *Parhyale* as an experimental system focus on developmental and evolutionary studies, much of it in the early embryo, and behavioural work of any kind is scarce. The next two chapters will cover attempts to establish its suitability to circadian research specifically. This chapter will discuss the development of a system to record the activity patterns of the amphipod and assess them for evidence of rhythmicity and endogenous control.

4.1.4.1 The fly approach...

With *Drosophila*, locomotor activity is typically monitored using the *Drosophila* Activity Monitoring (DAM) system (TriKinetics Inc). A standard DAM is equipped with 32 channels with a single infra-red (IR) emitter and receiver positioned to cross each one. Individual flies are placed into a glass tube stoppered with food at one end and cotton wool at the other, the tubes are loaded into the DAM and activity is registered and recorded by a data logger each time the IR beam is broken as the fly moves end to end in the tube (Figure 4.2). Monitors can be placed in temperature and light controlled incubators and subject to set photoperiods (light:dark, LD) for entrainment before shifting to continuous darkness (DD) to monitor endogenous rhythms. A larger version of the DAM has been used to monitor the swimming activity of *Eurydice pulchra* (Wilcockson *et al.*, 2011; Zhang *et al.*, 2013).
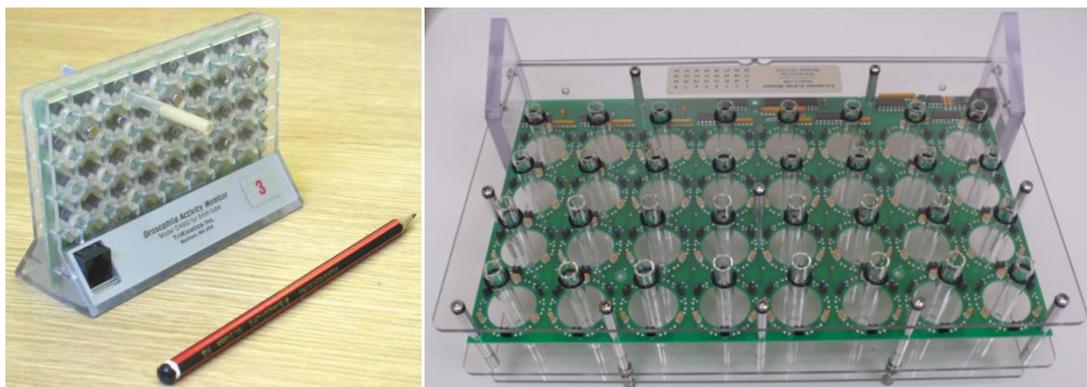
Figure 4.2: Left - *Drosophila* Activity Monitor with glass tube loaded into one channel. Right – Locomotor Activity Monitor with stabilised open-ended tubes, used to monitor *Eurydice pulchra* swimming activity. Both manufactured by TriKinetics Inc.

Data is typically grouped into 30 minutes bins and used to generate actograms depicting activity rhythms (Figure 4.3). It is also statistically analysed for periodicity using CLEAN spectral analysis and autocorrelation (Rosato and Kyriacou, 2006). The CLEAN algorithm is used to identify rhythmic components through a Monte Carlo process by which the data is shuffled and reanalysed with CLEAN 100 times to calculate the significance of the identified peaks, generating a plot with 95% and 99% confidence limits across the period domain – the highest peak passing the 99% confidence limit is taken as the fly's period. For autocorrelation, the data in *n* bins is compared to sequentially shifted copies of itself *n*/2 times with the correlation coefficient between data points calculated in each comparison and plotted, the biggest peak considered to represent the fly's period. Spectral analysis is the more powerful of the two approaches and is typically used in subsequent experimental analyses, while autocorrelation is generally run to support the former or make a decision on whether an individual is rhythmic or not – for example a period that crosses only the 95% confidence limit with CLEAN but is well supported by autocorrelation results could be considered rhythmic. A circadian periodicity in *Drosophila* is considered to be one that falls between 18 and 33 hours, while flies are considered arrhythmic if neither CLEAN nor autocorrelation show a significant periodic component, or lack a single dominant peak.

Using this approach one can identify the effects of mutations, temperature, entrainment photoperiod or other interventions on the organism. By collecting data for wild type flies and for mutants under the same conditions, or wild type under differing conditions, one can compare the mean period of each group to ascertain if the intervention has impacted the endogenous rhythm.
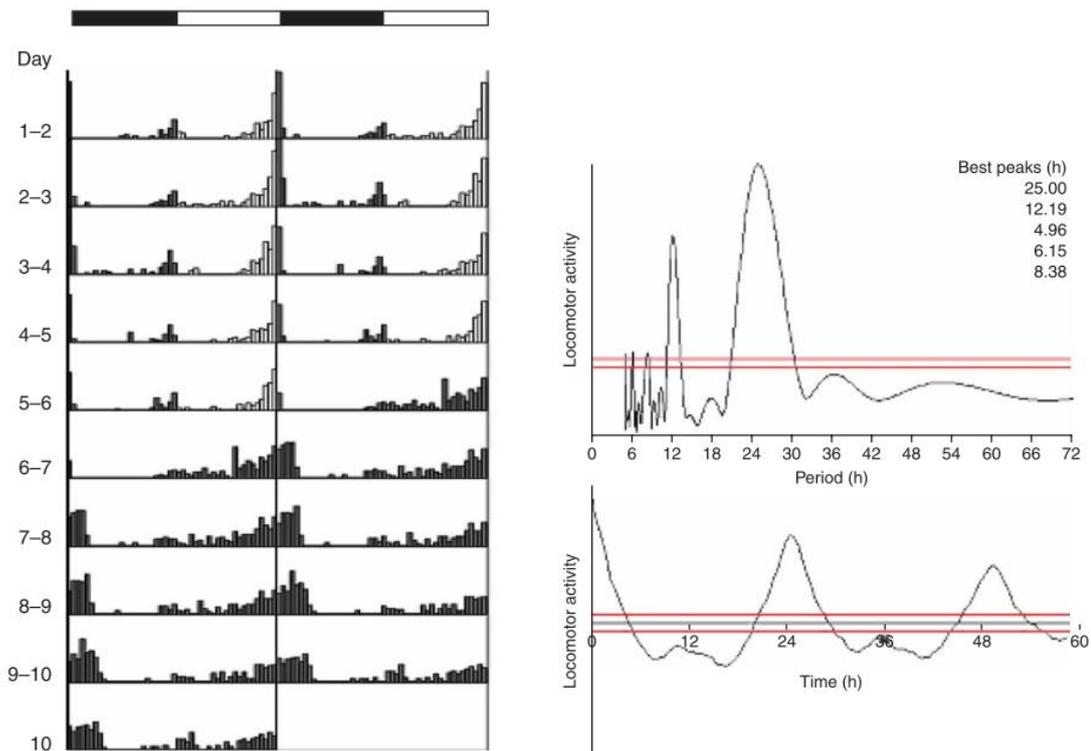
Figure 4.3: Locomotor activity data for a single fly in 5 days 12:12 LD, 5 days DD. Left – a double-plotted actogram depicting raw activity data per 30 minute bin. Across the top the black bar indicates lights off, the white bar lights on. Lights were permanently off from day 6. Right, top – CLEAN spectral analysis showing details of significant peaks crossing the 99% confidence limit. Right, bottom – autocorrelation analysis. This particular fly would be considered rhythmic, with an endogenous period of 25 hours. From Rosato and Kyriacou (2006)

4.1.4.2 ...also known as the failed *Parhyale* approach

Following the work on *Eurydice pulchra*, use of the larger Locomotor Activity Monitor (LAM) was the first consideration for monitoring *Parhyale*'s activity. This did not prove successful – while the tubes used for *Eurydice* were large enough for *Parhyale* to fit in they were clearly unable to move freely or turn easily, and it would be unlikely to record anything like natural activity under these conditions. An attempt was made

using even larger tubes, but even using stabilising rubber O-rings at the point of contact with the monitor this was hampered by the IR beams being broken by small water disturbances through the background vibrations of the incubator. This resulted in a strong tendency to suggest the animals showed constant low activity during the lights on phase, even when dead. As may be evident from the conclusion of the previous sentence, survival was also an issue – whether by increasing salinity through evaporation (the incubator was set at 25°C), depleted oxygen or sheer bad luck, assays of this type suffered unacceptably high mortality rates.

Finally, observations suggested that a vertical swimming assay may not be a good fit for *Parhyale*, whose activity tends to favour the horizontal plane in contrast to the beautifully clear substrate-to-surface-and-back behaviour of *Eurydice* (see http://www.eurekalert.org/multimedia/pub/62107.php for a video of this). *Parhyale* generally moves from substrate to substrate; occasionally it will rise in the water column but not actively return, instead simply allowing itself to sink.

After this first foray, it was decided to take a different approach. A system was required that could monitor a wider arena in which *Parhyale* might behave more naturally, could monitor from a top-down view to better detect movement in the horizontal plane, and would be more tuneable than the 1/0 logging of an IR beam break and so allow for a certain amount of noise. The rest of this chapter describes the development and use of an IR camera-based system to monitor *Parhyale* activity and collect data that can be used for analyses akin to that of *Drosophila*.

4.2   Methods

4.2.1   Husbandry

4.2.1.1 Founding

The starting population of *Parhyale hawaiensis* was provided by Ana Patricia Ramos of the Averof lab based at the Institut de Génomique Fonctionnelle de Lyon. These were themselves ultimately derived from the founding isolate described in the introduction to this chapter.

4.2.1.2 Husbandry

Parhyale were maintained in polypropylene containers, typically 25 x 20 x 9 cm with a 3 L capacity, with the lids used to cover 75% of the culture to slow evaporation. Artificial seawater (AS) with a specific gravity of ~1.022 or 30 ‰ was prepared using 30 g Tetra Marine Sea Salt per litre of distilled tap-water. Each container was filled with 1 L AS which was changed completely every two weeks using a sieve of mesh width 300 μm to prevent the loss of animals with waste water. A week after a water change, distilled tap-water was added to cultures to restore the initial salinity level. Each culture was aerated using an aquarium pump and airstone. A layer of crushed coral approximately 1 cm deep was provided as a substrate and to buffer pH. Cultures were maintained on a 12:12 light:dark cycle, illumination coming from fluorescent room lighting at ~300 lux, at ~20°C.

Animals were fed twice weekly, once with shredded carrot and once with a pinch of standard fish flakes. Carrot was stored, shredded, in a freezer in ice cube trays; a single cube of defrosted carrot was used to feed two - three cultures.

4.2.1.3 Experimental preparation

Prior to locomotor behavioural assays AS was prepared as above and aerated for a minimum of 24 hours using an aquarium pump and airstone. Roughly double the number of animals required for the upcoming assay were taken from a main culture using a 560 μm sieve and placed in a smaller container maintained to similar conditions with the exception of the crushed coral substrate, which was left sparser to enable individuals to be caught more easily. The experimental animals were fed *ad*

libitum in this container for one - two days. Animals used in experiments were caught using a plastic Pasteur pipette trimmed to widen the mouth to around 4 mm to avoid damage.

4.2.2   Locomotor behaviour

4.2.2.1 Hardware, software and analysis

Three TeckNet C016 USB webcams were modified to allow an image to be formed using infrared light through the removal of the infrared filter situated behind the lens. These were used in conjunction with iSpy v6.4.2.0 (https://www.ispyconnect.com/), an open source motion detection software package, running on a Lenovo G580 laptop with a 2.4GHz i3-3110M processor and 4 GB of RAM.

To monitor three or fewer groups/individuals the individual cameras were used as direct sources in iSpy. For four or more, each camera was instead connected first to the add-on software iSpyServer, typically used to allow remote access to cameras connected to other machines through the creation of a server on a local network; the camera is connected to iSpyServer and other machines running iSpy monitor the feeds remotely via the camera's assigned URL. Here, the feeds were instead linked as MJPEG video sources in iSpy running on the same machine by directing it towards the local host rather than the remote URL.

An iSpyServer feed can be linked repeatedly in iSpy, each time treated as an entirely new, independent video source that can have settings adjusted with no effect on others from the same feed. As iSpy allows the user to specify subsections of the feed to monitor for movement, this allows a single USB webcam to be used to monitor multiple separate static zones (Figure 4.4) with the only limit being the specifications of the hardware. The laptop used in these experiments was capable of running 18 video sources (6 per webcam feed) in this way at QVGA resolution, 5 fps, with minimal RAM impact and processor usage at approximately 50%.
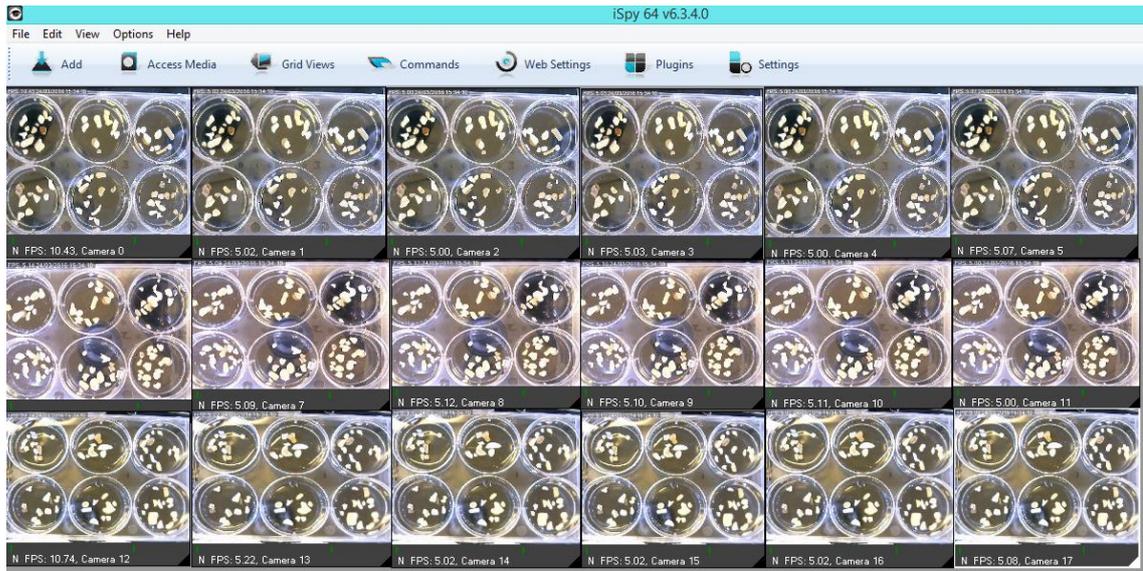
Figure 4.4: Top – multiple video sources in iSpy derived from a single USB webcam. Each row of cameras uses the same feed as their source. Bottom – a detection zone can be defined within each video source such that only motion within that zone will be registered and recorded by that source.

Locomotor experiments typically used the two frame motion detection setting with a trigger range of 2 – 70 and a minimum delay of two seconds (a 'refractory period' after recording an image). With these settings, for example, over the duration of 10 seconds of continuous movement in a detection zone, strong enough to surpass the minimum threshold of two but not the maximum sensitivity threshold of 70, five JPEG images would be captured and stored in the folder for that zone. Locomotor activity was determined by the number of images captured, typically grouped into 30 minute bins. Image file names were set by the time and date of their capture, allowing the names to be employed in grouping the images in the subsequent analysis. The file paths of the

images in each folder were output to a text file which was then processed using Microsoft Access 2007 to clip out the time and date each image was created and group them into 30 minute bins. The data was subsequently analysed for rhythmicity using the CLEAN spectral analysis algorithm and autocorrelation functions in BeFly! (Green, 2010).

Assessment of rhythmicity used a modified definition from Rosato and Kyriacou (2006). An animal was considered rhythmic if CLEAN showed a) at least one clearly significant peak under 33 hours, or b) barely significant peaks that were additionally supported by autocorrelation. The logic for the expansion of the lower bound in the first requirement is that crustaceans have shown circatidal and circadian rhythms that share underlying molecular components (Zhang *et al.*, 2013), and ultradian rhythms are also common. The same may prove true in *Parhyale* – this is therefore a broader assessment of locomotor rhythmicity than that typically conducted on *Drosophila*.

4.2.2.2 Group studies

A mixed group of six adults was selected from a prepared sub-culture (see 4.2.1.3) and placed in a glass crystallising dish 10 cm in diameter in 250 mL of AS. A small amount of crushed coral was spread sparsely across the base – a complete covering was not feasible as the animals were almost undetectable against a background of coral under infrared illumination. The dish was placed on a black plastic sheet to further enhance the contrast between *Parhyale* and the background. The dish was covered with a petri dish lid held in place with blu tac to reduce evaporation (Figure 4.5).
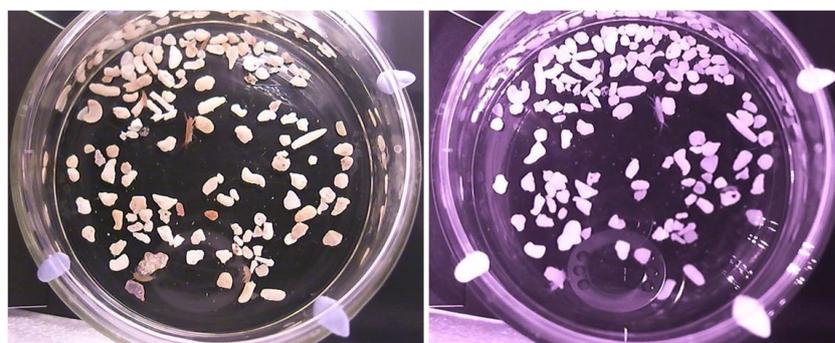
Figure 4.5: Arena set-up for group experiments. Left - light phase. Right – dark phase under infrared illumination.

A webcam was suspended above the dish using a clamp and stand. Light was provided using a fluorescent strip light unit from a *Drosophila melanogaster* light box unit (~300 lux) constructed by the Mechanical Workshop of the University of Leicester's College of Medicine, Biological Sciences and Psychology, and controlled with a digital plug-in timer. Infrared illumination was provided by a Foscam FI8910W IP camera, which has a light-sensitive auto-on IR function. For two full days after transfer to the experimental arena the animals were subject to a 12:12 LD cycle. They were subsequently monitored for five days in continuous darkness before the data was collected.

This process was conducted for three mixed adult groups of six, at 20°C. After completion of the DD phase the water in the dishes was changed, food was supplied and the animals allowed to reentrain to a 12:12 photoperiod for two days before the individuals were transferred to separate wells of a six well cell culture plate for further assay (see below). The temperature during experiments was monitored using a HOBO data logger (Onset Computer Corporation).

4.2.2.3 Individuals

Adults of both sexes were used in experiments monitoring isolated animals, some of which came from groups that had already been assayed (see above), while others were selected directly from the prepared subculture. Individuals were placed in the wells of plastic six well cell culture plates with 10 mL of aerated AS and a small amount of crushed coral. The plate lid was used to slow evaporation. Up to three plates containing a total of 18 animals were monitored in a single experiment using the iSpyServer method (see 4.2.2.1), each well monitored on a separate linked feed; in some experiments wells were left empty as a negative control to ascertain the reliability of the motion detection algorithm. Experiments were conducted using two – four days 12:12 LD followed by five days of DD, at 20°C. Later instances used dimmable LED light strips as the light source due to technical issues.

A vertical assay was also trialled. This entailed the use of clear plastic cylinders 25 mm in diameter and 95 mm in height as containers with crushed coral to a depth of 0.5 cm and 25 mL of aerated AS. Evaporation was prevented with the use of cellulose acetate plugs. The motion detection zones for each cylinder monitored were set 1 cm above

the coral substrate so as to only detect the amphipods during vertical swimming behaviour (Figure 4.6). This experiment was conducted once on 8 adult males, using the fluorescent strip light source during the LD phase, at 20°C.

Figure 4.6: Vertical swimming assay. Transparent rectangle depicts motion detection zone.

4.3 Results

4.3.1 Activity profile in LD

All animals exhibited nocturnal behaviour, showing higher levels of activity during the dark phase of LD. Activity would notably rise immediately after lights off and persist into the early light phase before dropping to very low levels until the beginning of the next dark phase (Figure 4.7).
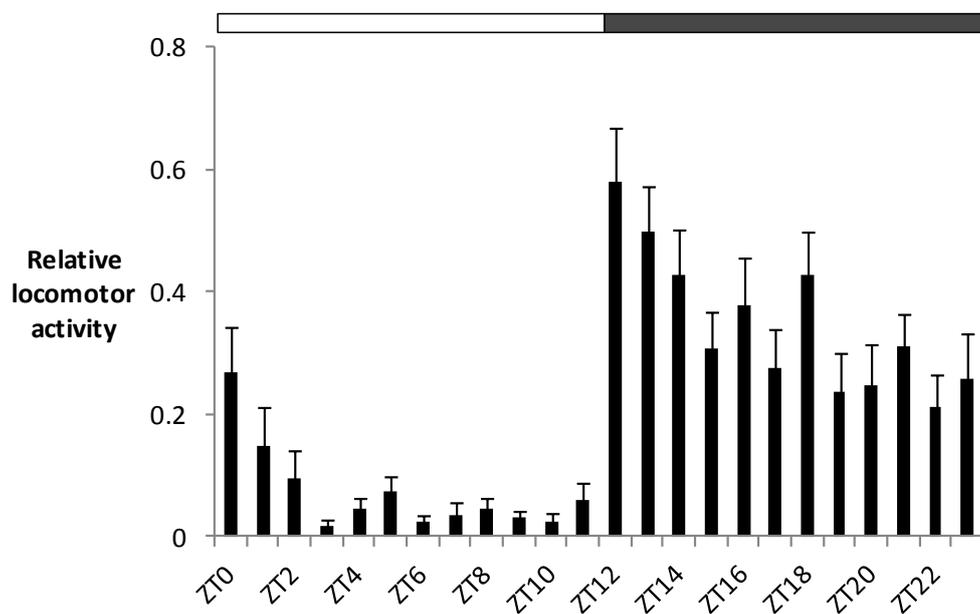
Figure 4.7: Mean normalised activity (+ S.E.M.) per hourly bin under LD 12:12 conditions (n = 18). For each individual, activity in each bin was normalised to the highest value across the complete time series. ZT – *zeitgeber* time.

This finding aside, there was no obvious pattern to activity – individuals were highly variable in where their nocturnal peak of activity came, whether or not they showed a burst of activity at lights-on, and whether they were active at all during lights-on.

4.3.2 Detected periodicities

4.3.2.1 Groups

Of the three group assays Group 1 showed evidence of an endogenous rhythm supported by both CLEAN and autocorrelation analysis (Figure 4.8), with a dominant

peak of 23.9 hours and a significant 12.85 hour component. While Group 2 showed a significant peak at 22.98 hours it was not clearly dominant and lacked autocorrelation support. Group 3 meanwhile showed two significant peaks, neither dominant, at 8.02 and 21.73 hours, again with poor autocorrelation.
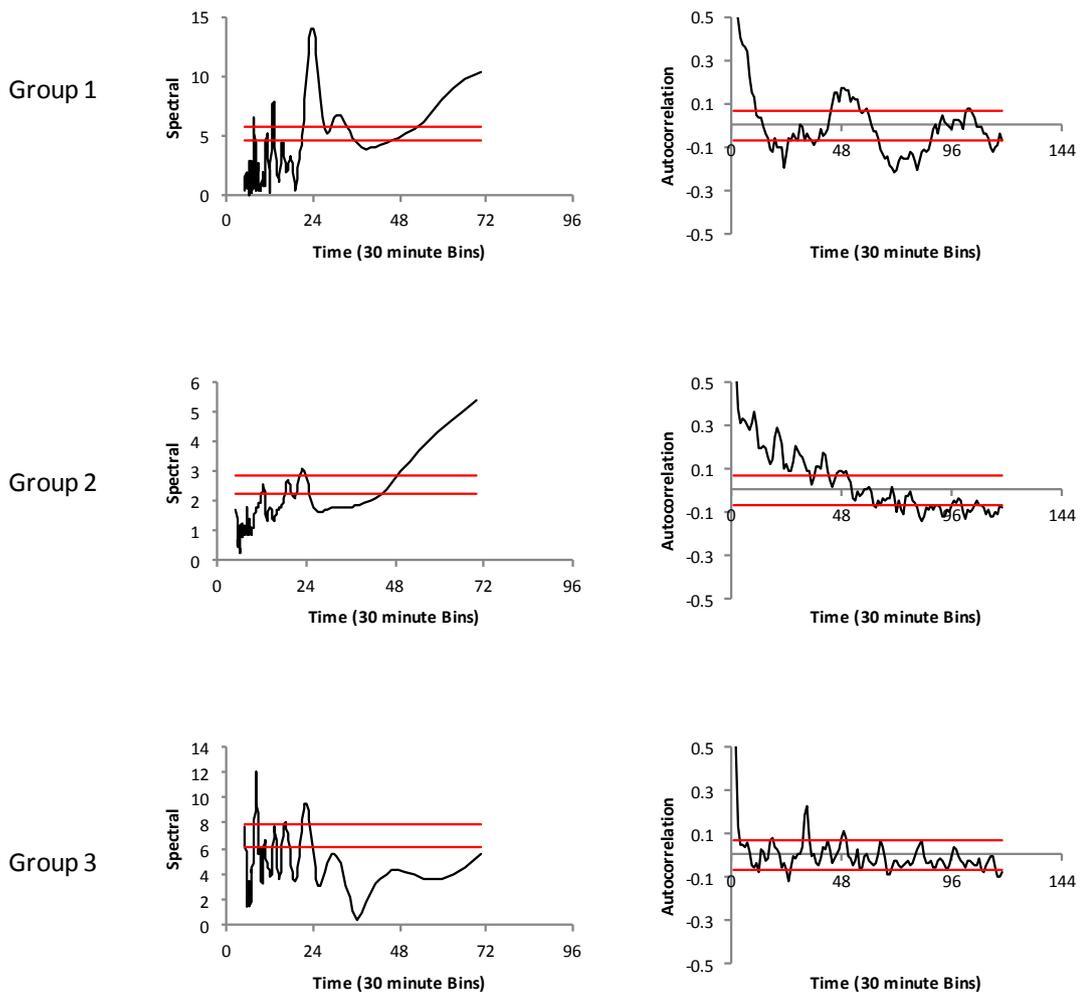
Figure 4.8: CLEAN spectral analysis and autocorrelation results from group assays. Each group consisted of adults of both sexes, entrained to LD 12:12 for two days before five days of DD, at 20°C. Results are calculated using only DD activity data.

4.3.2.2 Individuals

Data for a total of 68 individuals was collected from five successfully completed assays, with the data of five individuals found to have died before the end excluded from further analysis. By the criteria set out above 29 out of 63 individuals (46%) showed evidence of an endogenous locomotor activity rhythm (Table 4.1). Two components stood out when surveying these periods (Figure 4.9). Of the 29 rhythmic individuals, 17 showed a dominant periodic component of ~24 hours (defined as any falling between 18 – 33 hours), while eight showed a dominant component of ~12 hours (defined as any falling within 10 – 13 hours). Of the 25 animals showing at least one of these components, 10 showed both, with the ~24 hour period the dominant component in six cases, while the ~12 hour period was dominant in three cases. The mean of the ~24 hour components was 24.84 ± 2.28 S.D., while the mean of the 12 hour components was 11.95 ± 0.55.

Figure 4.10 shows CLEAN and autocorrelation results with accompanying double-plotted actogram for two representative individuals, one with a dominant ~24 hour period and one with a dominant ~12 hour period. As the activity patterns are quite noisy, the same data grouped in 3 hours bins is shown as a single line chart in Figure 4.11.

Table 4.1: Individual *Parhyale hawaiensis* showing evidence of endogenous control of locomotor activity based on CLEAN spectral analysis, autocorrelation or both. Red – a dominant periodic component outside the most common ~12 and ~24 hour ranges. Blue – in animals with both common periods, this highlights which was dominant.

Assay ID	Sex	CLEAN (99%)	Autocorrelation support	Period of dominant component	~12 hour period	~24 hour period
P002I	Male	Yes	No	25.43	12.19	25.43
P002I	Male	No	Yes	38.55	11.38	22.13
P002I	Female	Yes	Yes	27.16	-	27.16
P002I	Female	Yes	Yes	25.43	12.58	25.43
P002I	Male	Yes	Yes	12.32	12.32	-
P002I	Female	Yes	Yes	23.9	11.95	23.9
P003I	Male	Yes	No	23.94	-	23.94
P003I	Male	Yes	Yes	24.24	12.01	24.24
P003I	Male	Yes	No	22.53	12.12	22.53
P003I	Male	Yes	No	11.89	11.89	26.6
P003I	Male	Yes	Yes	29.02	-	29.02
P005I	Male	Yes	Yes	24.39	-	24.39
P005I	Female	Yes	Yes	6.36	-	-
P005I	Male	Yes	Yes	23.9	-	23.9
P005I	Female	Yes	Yes	8.92	-	22.98
P005I	Male	Yes	No	25.43	-	25.43
P005I	Female	Yes	Yes	11.27	11.27	-
P005I	Male	Yes	No	30.64	-	30.64
P005I	Male	Yes	No	23.43	-	23.43
P005I	Female	Yes	Yes	12.58	12.58	24.9
P005I	Male	Yes	Yes	11.95	11.95	-
P009I	Male	Yes	Yes	23.9	-	23.9
P009I	Female	Yes	No	17.57	-	-
P009I	Male	Yes	No	10.48	10.48	-
P010I	Male	Yes	No	11.83	11.83	20.96
P010I	Male	Yes	No	12.19	12.19	-
P010I	Male	Yes	Yes	26.56	12.45	26.56
P010I	Male	Yes	Yes	26.56	-	26.56
P010I	Female	Yes	Yes	22.55	-	22.55
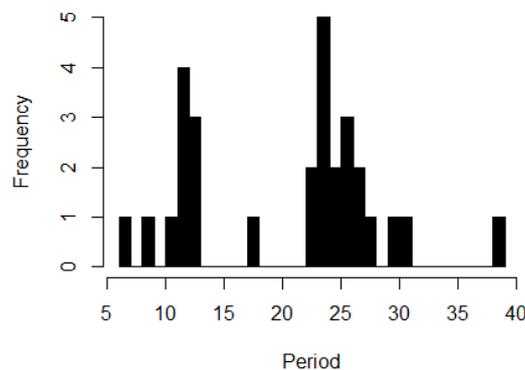
Figure 4.9: Distribution of dominant peak periods of individuals, identified using CLEAN spectral analysis (n = 29).
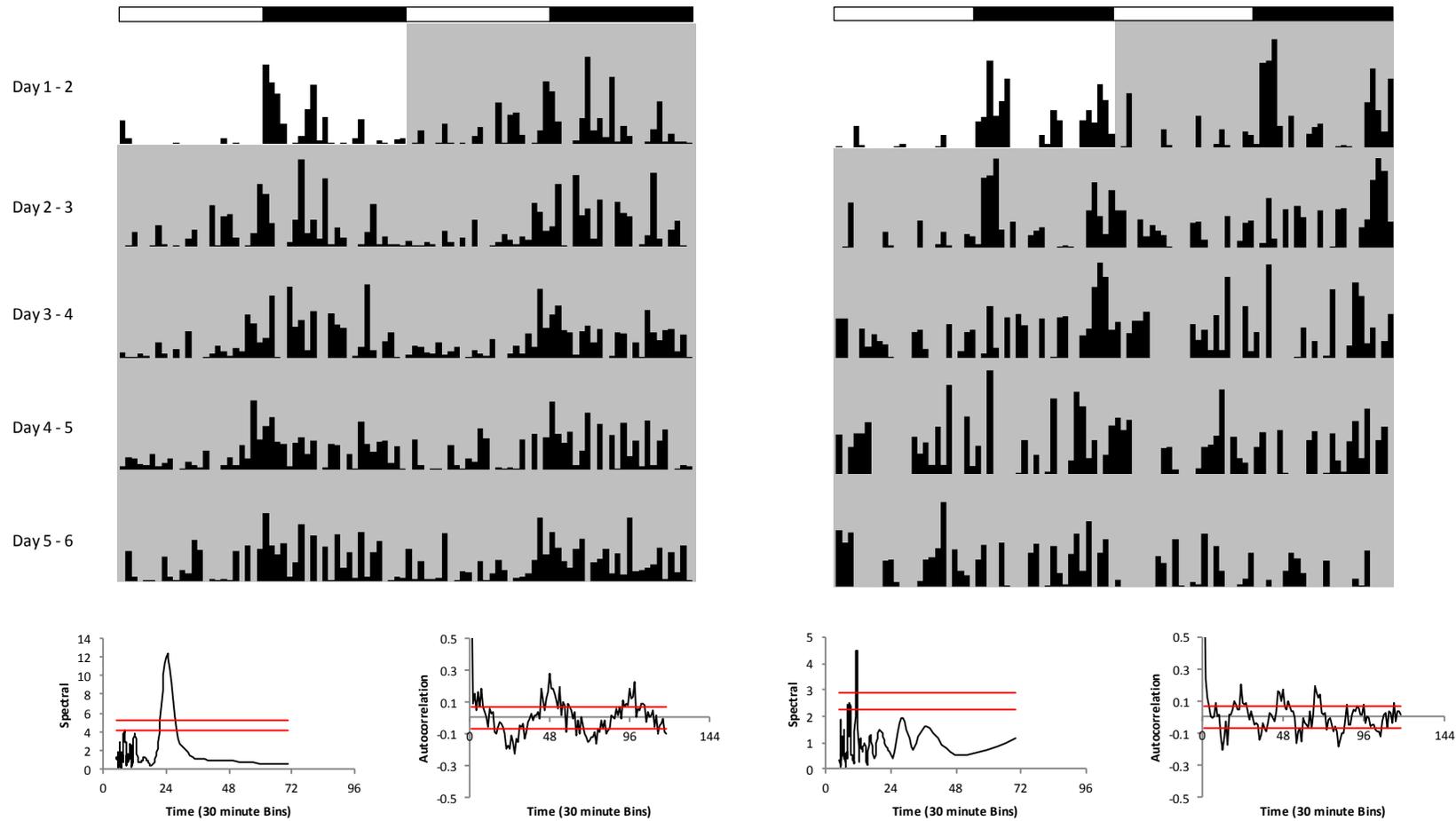
Figure 4.10: Double plotted actograms (top) and CLEAN/autocorrelation results (bottom) of representative individuals with dominant ~24 hour (left) and ~12 hour (right) periodic components. Activity is plotted as number of images recorded per 30 minute bin. The black and white bars represent the lights-on and lights-off phases of the prior entrainment period. Day 1 was in LD; every subsequent day in constant darkness (marked in grey overlay).
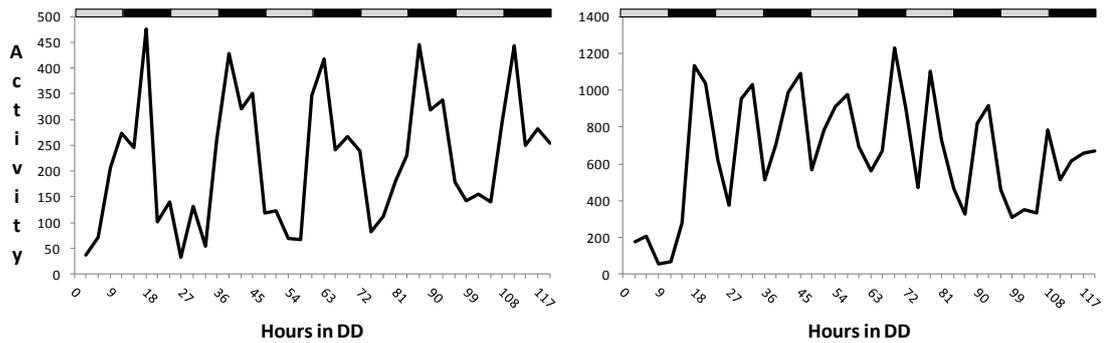
Figure 4.11: Locomotor activity of representative individuals with dominant ~24 hour (left) and ~12 hour (right) periodic components in five days of constant darkness, grouped into 3 hour bins. Activity is plotted as number of images recorded per 3 hour bin. The black and grey bars represent the lights-on and lights-off phases of the prior entrainment period.

4.3.3   A set of observations

4.3.3.1 Males and females

Of the 63 surviving individuals that were assessed for rhythmicity, 44 were male and 19 female. Of the 29 with dominant periodicities, 20 were male and 9 female. There was therefore no significant difference in the ratio of rhythmic to non-rhythmic individuals observed between the sexes (χ^2 = 0.0196, df = 1, p >> 0.1).

4.3.3.2 Groups and individuals

Three groups of six were tested for rhythmicity, the results depicted in Figure 4.8. The individuals from these tests were subsequently tested in isolation after a period of re-entrainment, in assay P005I. Of the 10 individuals in that assay that were found to be rhythmic (Table 4.1), five of them came from Group 1, which was the only group to generate robust statistical evidence of rhythmicity, with a dominant peak of 23.9 hours and a significant 12.85 hour component. Two of those five, furthermore, generated actograms in which rhythmicity was obvious to the eye before statistical testing.

4.3.3.3 Vertical assay

P003I was the vertical assay, with five of the eight males showing evidence of endogenous control of locomotor activity. Two of those five gained particularly robust support from both CLEAN spectral analysis and autocorrelation.

4.3.3.4 The emergence of the second peak

In certain individuals showing a ~12 hour component a second peak of activity emerged in constant darkness. While this can be seen in Figure 4.10 the rhythmicity of that individual is not particularly clear when looking at the actogram alone. Figure 4.12 depicts another individual that better illustrates this phenomenon.
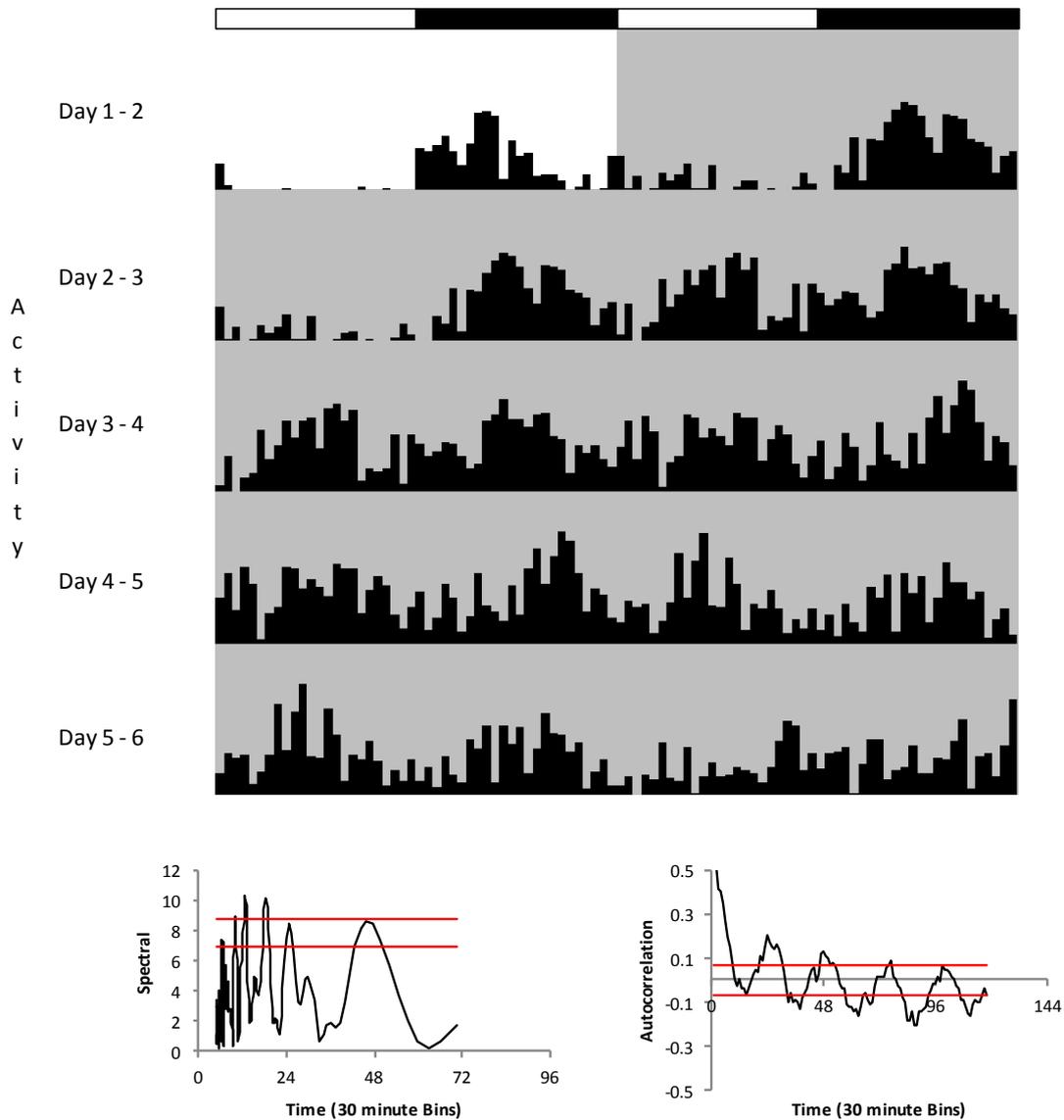
Figure 4.12: Actogram of individual depicting emergence of a second daily peak of activity during DD. The black and white bars represent the lights-on and lights-off phases of the prior entrainment period. Day 1 was in LD; every subsequent day in constant darkness (marked by grey overlay).

4.3.4 Temperature control

A temperature controlled room was used for these studies, allowing for the use of a relatively bulky set-up of cameras, infrared lighting and computer equipment. It did, however, prove less reliable than an incubator, the temperature prone to deviations of 1 - 2°C either side of 20°C and occasional minor spikes. Figure 4.13 shows the temperature during DD plotted against the activity of rhythmic individuals from two separate assays. CLEAN spectral analysis of this temperature data gave weak support for the possibility of circadian variation only in dataset B, with a broad, bimodal-like peak suggesting a periodic component of 23.43 hours. The locomotor activity plotted against this data in Figure 4.13 had a strong periodic component of 24.39 hours, and a Spearman's rank-order correlation indicated no significant correlation between temperature and activity level (r_s = -0.0757, p = 0.41).

Spectral analysis of the temperature data of a third assay similarly did not find any periodic components, and autocorrelation in all three cases gave no support to rhythmicity. Temperature readings for the remaining two assays are unavailable due to technical failure (and the failure to recognise the first technical failure as such is due to researcher failure). As these latter two assays were run during a time when the room's temperature control had been established within the parameters shown above the activity data was included in the results, but caution is warranted.
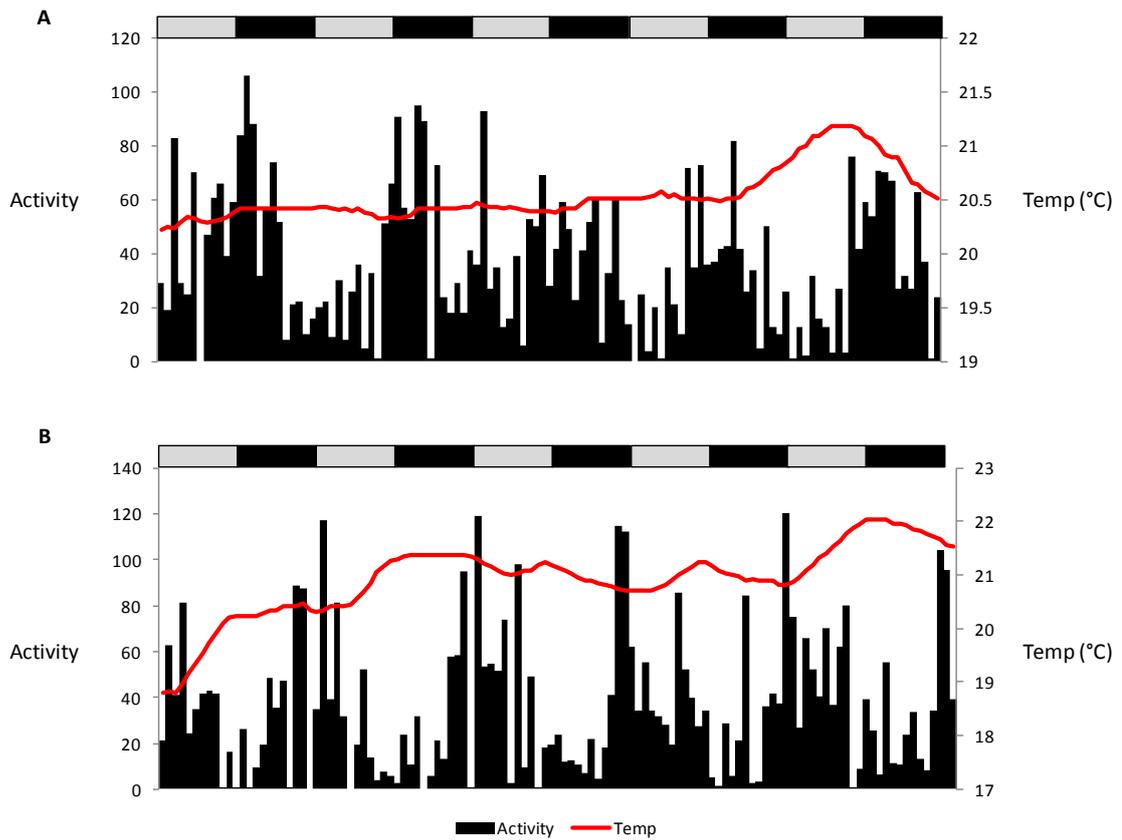
Figure 4.13: Activity levels of two rhythmic *Parhyale* plotted against temperature in DD in separate assays. A) shows a late temperature spike of < 1 °C, while B) shows a more variable temperature pattern from 18.8 – 22.04°C. CLEAN spectral analysis weakly supported a circadian periodicity to temperature variation in B) only. The black and grey bars represent the lights-on and lights-off phases of the prior entrainment period.

4.4 Discussion

4.4.1 Summary of results

The data suggest that the locomotor activity of *Parhyale hawaiensis* is under the control of the endogenous clock, with 46% of animals showing evidence of rhythmicity in constant darkness as determined by CLEAN spectral analysis and autocorrelation. The majority of dominant periodicities were classified as either ~12 hour or ~24 hour components. Rhythmic individuals were identified in assays monitoring both horizontal and vertical movement. The ratio of rhythmic to arrhythmic individuals is similar in both males and females, while the rhythmicity of groups may emerge from the number of rhythmic individuals it contains. Temperature control during the studies was not as robust as the ideal, but at worst fluctuations remained within 1 - 2°C of the set value and did not appear to occur rhythmically or drive activity patterns.

4.4.2 Evidence of endogenous rhythms

4.4.2.1 A nocturnal animal

One of the few behavioural studies on *Parhyale hawaiensis* focused on feeding preferences (Poovachiranon *et al.*, 1986). A daily feeding rhythm was identified through the ingenious method of collecting and weighing faecal pellets every two hours, having determined a typical period of 20 – 30 minutes between feeding and defecation. The results showed that in single sex adult groups both males and females showed a significant maximum feeding rate during the night hours, from 2 – 6 am, and a minimal feeding rate in the afternoon to evening, with the female minimum at midday while the male minimum between 4 – 6 pm (Figure 4.14). A group of mixed sex juveniles, meanwhile, showed no feeding rate pattern.
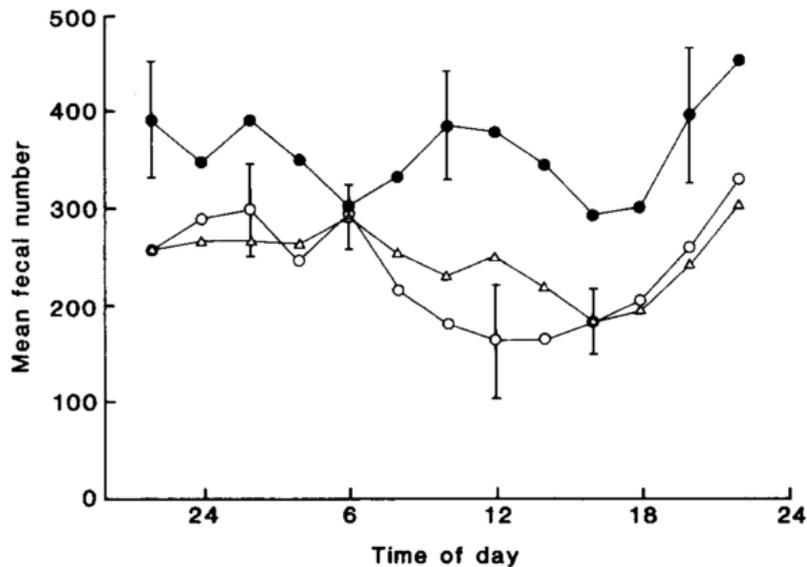
Figure 4.14: Feeding rates across 24 hours as measured by faecal production. Δ – male group. o – female group. • - mixed sex juvenile group. Error bars show typical standard deviations. From Poovachiranon *et al.* (1986).

The results described in this chapter support these findings – *Parhyale hawaiensis* gives clear indication that it is a nocturnal or photophobic animal, with activity greatly increased during the dark hours (Figure 4.7). Furthermore, as the feeding rate study did not investigate rates under constant conditions, we have generated the first evidence that this nocturnal pattern may be under the control of the endogenous clock.

4.4.2.2 Twin peaks

Two common periodic components were identified, ~12 hours and ~24 hours. While the latter can with some justification be termed a circadian rhythm (although see 4.4.3.3), it is less certain what to make of the other component. It may be that *Parhyale* possesses a bimodal circadian activity pattern – the immediate peak of activity after lights-off was nearly universal, but while for some animals activity persisted uninterrupted throughout the dark phase, or steadily dropped off without reprise, in others a lull in activity was seen followed by a rise towards lights-on. Unfortunately the noisiness of the general pattern and the inconsistency with which such patterns were seen even in one animal across days make it difficult to relate such LD activity to subsequent DD rhythms. Possibly relevant is the data from an early group

study that is not included in the results due to unacceptably poor temperature control, the data gained when temperature was in theory set at 18°C but fluctuated between 15.5 – 20.7°C in a circadian manner. In this group, a pattern was repeated across two days of LD in which both lights-on and lights-off peaks of activity were seen, which persisted over five days of DD, generating highly significant peaks in spectral analysis and strong autocorrelation support (Figure 4.15).
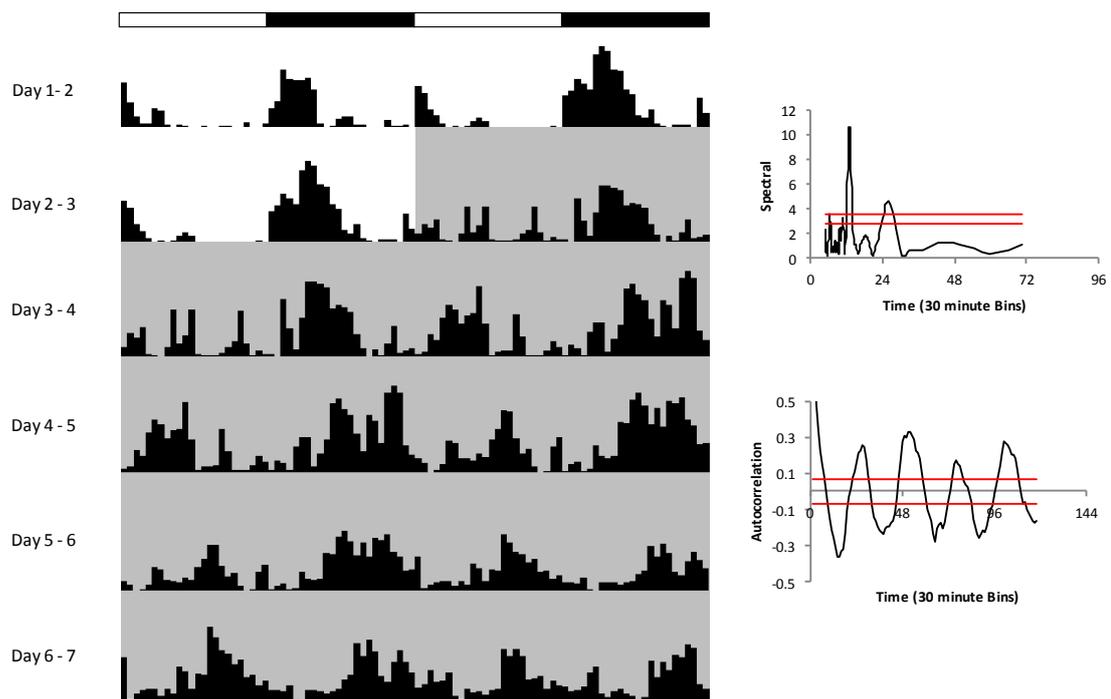
Figure 4.15: Double-plotted actogram and CLEAN spectral analysis and autocorrelation for early group study. Day 1 and Day 2 were in LD, all other days in DD (marked by grey overlay). Note the lights-on peak that exists in LD and persists, expanded, in DD. Significant peaks were 12.71 and 25.98 hours.

While the effect of temperature cannot be discounted in this case, it does suggest the possibility of a bimodal activity pattern present in at least some animals, and this lights-on/lights-off pattern of activity peaks reflects that found in the crayfish *Procambarus clarkii* (Page and Larimer, 1972). One could further speculate that the fact that this lights-on peak was not seen in isolated animals yet emerged in some cases during DD (Figure 4.12) indicates a social component to activity patterns, the presence of conspecifics perhaps required for activity to be seen during light hours – given the otherwise nocturnal behaviour and the high population density of these

animals in the wild (Poovachiranon *et al.*, 1986) this would not be surprising. This assay, incidentally, inspired the idea that *Parhyale* may be rhythmic only in groups, as at that point studies of individuals – conducted one or two at a time as the camera system had not been fully developed – suggested no persistence of activity rhythms in DD. Under continuous dim red illumination the amphipod *Talitrus saltator* shows reduced drift in activity peaks in groups compared to isolated individuals (Bregazzi and Naylor, 1972) indicating a social component to its rhythmicity, and it was considered possible that a more extreme example of the phenomenon was present in *Parhyale*. However from the small amount of data communicated here in which not all groups showed robust rhythmicity, and those that did proved to have a large number of rhythmic individuals, it seems likely that this early absence was due to the relatively low percentage of rhythmic individuals in the total population. Group rhythms may instead depend on the rhythms of their component parts rather than emerge as a social phenomenon.

A second thought regarding the ~12 hour component is that it may represent a circa-tidal rhythm. The habitat of wild *Parhyale* would certainly be conducive to the evolution of such an endogenous system – found in intertidal mangrove zones and so subject to tidal fluctuations in salinity, oxygen levels and turbidity, the ability to entrain to such a cycle would likely bring strong benefits in terms of control of physiology and biochemistry, as well as feeding behaviour and predator avoidance. The animals used in these experiments, however, have been lab animals for nearly two decades, and this after an indeterminate time spent living on the filtration system of a large scale aquarium. They have not been subject to tidal entrainment for hundreds of generations, and given that wild-caught *Eurydice pulchra*, given a substrate to hide in, will not display circa-tidal swimming without entrainment by agitation (Jones and Naylor, 1970), it is rather optimistic to suggest that *Parhyale* has managed to maintain such a rhythm down the long years of Tupperware existence. That said, if the capability exists and if it shares molecular components with the circadian system, it might show glimpses of itself as part of a wider response to light entrainment, particularly in 12:12 conditions. Further work is necessary to investigate this aspect of *Parhyale*'s locomotor activity.

4.4.2.3 A noisy pattern

While the activity may be under the control of the clock, it is a distinctly noisy pattern.

The circa-tidal swimming of *Eurydice pulchra* is a neat and defined phenomenon, while

in *Parhyale* even robustly rhythmic individuals, as identified through spectral analysis,

tend to generate messy actograms (Figure 4.16)
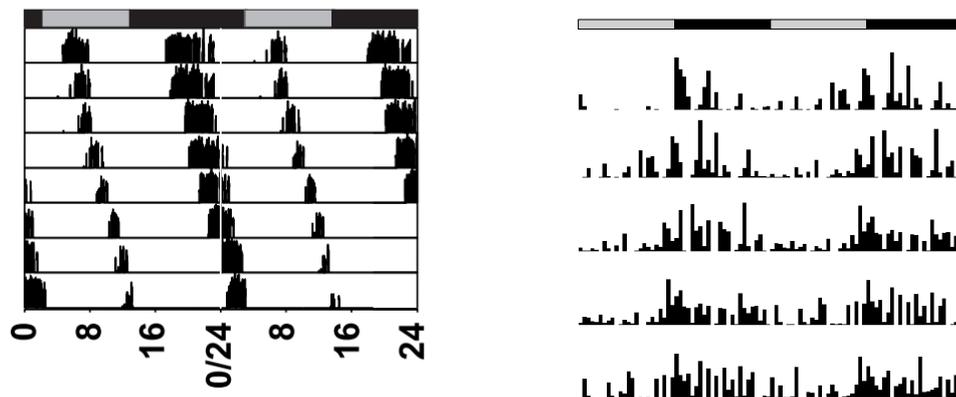
Figure 4.16: Comparison of circa-tidal swimming behaviour in *Eurydice pulchra* (left) and circadian
locomotor activity in *Parhyale hawaiensis* (right) in constant conditions on double-plotted 24 hour
actograms. *E. pulchra* image from Zhang *et al.* (2013).

Similarly messy results have been obtained from the lobster *Nephrops norvegicus*

(Aguzzi and Sardá, 2007) and *Procambarus clarkii* (Page and Larimer, 1972) and it is

perhaps an inevitable consequence of studying general locomotor activity rather than

a specific behaviour.

4.4.2.4 No difference between the sexes

With no prior knowledge of rhythmicity in this species, it was considered sensible to

assay both male and female individuals in case one sex might prove more robustly

rhythmic than the other. Differences in circadian activity and the percentage of

rhythmic individuals have been identified in the parasitoid wasp *Nasonia vitripennis*

(Bertossa *et al.*, 2013), for example. The inclusion of females brought complications in

that it was difficult to tell without microscopic inspection if they were carrying

fertilised eggs, and indeed two individuals were found to have produced offspring at

the end of the assay. These were too small to register with the motion detection

software and the identified periodicities of these individuals were no different from

females that had not released offspring and so their data was retained for analysis, but

this is an issue that should be borne in mind both when considering the results and future assays. It cannot be discounted that both the mated or unmated status of the female and the presence or absence of offspring might affect activity patterns – mating activity is under control of the clock in *Drosophila* and driven by female behaviour (Sakai and Ishida, 2001), for example.

4.4.2.5 Vertical swimming behaviour

Five out of eight adult males showed a rhythm in vertical swimming behaviour, in which the camera was set only to detect motion that occurred 1 cm or higher above the substrate. The periods identified did not differ from those obtained from top-down assays, and only two of the five were robustly supported results – as with the top-down assays, actograms were noisy and it is likely that the detected movement was general swimming behaviour that went upwards because of restriction in the horizontal plane, rather than deliberate substrate-to-surface activity.

4.4.2.6 Petering out

A number of animals that did not return significant results in spectral analysis or autocorrelation did in fact show a persistent rhythm in DD for one or two days, before either dropping away to very low levels of activity or showing arrhythmic patterns – the results above show only those animals maintaining rhythmicity over five days in constant conditions. On this basis it is suspected that the percentage of rhythmic animals in a population may be higher than that described here, should improvements in the experimental set-up and assay be applied. This is addressed in more detail in the next section.

4.4.3 Future work

It seems reasonable to conclude that the perfect assay of *Parhyale* locomotor activity has not yet been devised and conducted, but the experience gained in obtaining these results can inform subsequent refinements. Such improvements will increase the power of the assay to identify significant differences between experimental conditions or genetic lines without requiring huge numbers of participants.

4.4.3.1 Preparation

If one wished to be kind to the process used to select individuals for these assays it could be described as 'randomised', which is of course a key aspect of good science. In all honesty, however, selecting animals by taking a scoop of gravel and choosing adults hiding within on the basis of whether or not they can escape the pipette does not make for a particularly well-controlled study. The age of the individuals, their nutritional status and their moult status are not taken into account by this method. The moult phase in particular is a stressful and vulnerable period that varies in frequency from around every 2 – 4 weeks depending on age, nutritional status and population density (Konstantinides and Averof, 2014), and typical locomotor activity is unlikely to be seen in while it is ongoing and perhaps afterwards given that no food is supplied during the assay. Regarding age, it has been shown that *Drosophila* become less rhythmic as they grow older (Grotewiel *et al.*, 2005) and if this is true in *Parhyale* then failure to use young adults will impact on the success of the assay.

Given *Parhyale*'s mating system, however, this can improved. Mate pairs in amplexus can be isolated until they detach, after which fertilised eggs can be removed from the female's marsupium and raised to adulthood, allowing assays to be performed on an age cohort, and with a more watchful eye on moult timing. Such an approach will furthermore give more control over genetic variation, and these changes will make it more similar to the study of locomotor behaviour in *Drosophila*, in which virgin females and males are mated and the offspring used for experimentation. This approach would also solve the issue of assaying the activity of females without knowing their mated status.

4.4.3.2 Assay

In these experiments the entrainment period prior to the start of DD lasted two – four days, while DD lasted five days. Ideally both phases would be longer, as a minimum of five days entrainment is standard for *Drosophila* assays (Rosato and Kyriacou, 2006), and longer assessment of DD activity strengthens the statistical analysis. While *Parhyale* appears to be capable of surviving long periods of starvation, after a week activity can drop to very low levels in some individuals. This pattern is sometimes disrupted by sudden bursts of high activity seen in animals subsequently found dead,

suggesting energy conservation followed by a terminal desperation to locate food. Obviously this would not be an issue if food were provided during the assay as it is with *Drosophila*, and this was attempted without success – provision of even minimal amounts of food seemed to quickly spoil the small volume of water each animal was in and very high mortality was observed. The refusal to suffer starvation gladly was also seen in their cannibalistic tendencies, as attempts made to observe the activity patterns of mate pairs over the usual 7 day span mostly ended with a surviving male surrounded by sadly identifiable debris.

To improve the situation it will be necessary to design an apparatus that can accommodate food provision without spoilage and oxygen depletion, so that individuals are maintained in conditions closer to their day to day culture. Initial thoughts on this subject tend towards a tube sealed at the base with mesh wide enough to allow faeces and food detritus to pass while keeping the animals inside, suspended in a larger body of water that is aerated with an aquarium pump. In this way a reasonable amount of food can be provided while the water is constantly replenished with oxygen and waste passes away from the experimental animals. Additionally by using tubes set within the water reservoir and breaching the surface any disturbances from the aeration process will not impact on the area being monitored.

Once this issue has been resolved, more complex assays will be possible, such as determining if the feeding rhythm described by Poovachiranon, Boto and Duke (1986) is under endogenous control. It will also be of interest to attempt entrainment by other potential *zeitgebers* including agitation – which can be used to investigate putative circa-tidal rhythms – and food provision, as seen with the cavefish *Phreatichthys andruzzii* (Cavallari *et al.*, 2011).

Finally, the temperature must be better controlled so that we can be certain that whatever we observe is not following temperature fluctuations. The use of an incubator would be the ideal solution but as mentioned previously the set-up as it stands is rather bulky. It may yet prove that a larger version of the *Drosophila* Activity Monitoring system, using infra-red sensors connected to a remote data logger, is the

best approach – this will of course require addressing issues of mortality rates and the animals' freedom of movement.

4.4.3.3 A *circadian* rhythm?

Some of the results detailed here have been suggested to represent a circadian rhythm, on the basis of an approximately 24 hour period persisting in constant conditions. But this meets only one of the three standard criteria for determining if a cycle of behaviour is controlled by the endogenous circadian clock (Rosato *et al.*, 1997). It would take a considerable amount of chutzpah to attempt to make a virtue of the difficulties in maintaining a constant 20°C and claim that temperature compensation has been shown here, and so this shall be filed – along with resetting via entrainment – as work yet to be conducted. While attempts at both were considered during this research it was ultimately decided that it was more important to gather basic data and try to improve the assay quality, so that subsequent experimentation might obtain the necessary results more easily.

4.4.3.4 Genetic tinkering

Looking further into the future, once the assay has been optimised the considerable genetic resources available for *Parhyale* can be marshalled to investigate the effects of cell ablation, RNAi knockdown and gene mis/over-expression, and generate mutant lines for comparison to wild type. Achieving this, of course, first requires the identification of genes of interest. This is the subject of the next chapter.

4.4.4 Conclusions

This chapter relates the first known evidence of endogenous locomotor rhythms in the model organism *Parhyale hawaiensis* in constant darkness after entrainment in LD 12:12. An infra-red camera system has been designed and employed to gather activity data indicating a circadian period of 24.84 ± 2.28 S.D., with an approximate 12 hour component (11.95 ± 0.55) also commonly present. Activity is at its highest during the dark phase of entrainment, which persists in DD and suggests *Parhyale* to be a nocturnal or photophobic species, in line with published feeding timing data. Further work is required to optimise the behavioural assay and determine if the cycle exhibits temperature compensation and light-mediated resetting. *Parhyale* is a model organism

with recent successes in CRISPR/Cas9 targeted mutagenesis, opening the door to the study of circadian mutants in the style of the classic *Drosophila* literature. An important step towards this will be the identification of the genes presumed to underlie the endogenous clock in this species.

Chapter 5   The molecular clock of *Parhyale hawaiensis:* where is it?

5.1   Introduction

5.1.1   A familiar tune

A lengthy introduction to this chapter will not be necessary. Chapter Four describes the first evidence of endogenous control of locomotor activity patterns in the amphipod *Parhyale hawaiensis*. This behavioural work requires further refinement and expansion, but once a reliable assay has been established *Parhyale'*s genetic toolkit provides the potential for sophisticated dissection of the molecular system that is presumed to underlie any observed rhythms. Of course to dissect a molecular system it must first be identified, and so once again we find ourselves searching for genes.

5.1.2   Aims and objectives

Given the successful use of the RNA-seq approach in identifying genes of interest in *Euphausia superba* (Chapter Three), it seemed reasonable to employ the same technique with *Parhyale*.

The *Euphausia superba* transcriptome was based on mRNA from head tissue, as a number of circadian oscillators are found here in other crustaceans - such as in the retina, the eyestalk and the brain (Strauss and Dircksen, 2010) - and so this seemed like the most promising tissue to use when searching for circadian genes. At the point at which the work detailed below was undertaken a number of transcriptomic resources already existed for *Parhyale*, but these were derived from ovarian and embryonic tissue and searches produced little evidence of orthologs of the canonical circadian genes aside from a possible small fragment of *bmal1*. This chapter therefore describes the creation of a head transcriptome for *Parhyale hawaiensis* and its subsequent mining for circadian genes. Very recently a draft genome has been produced for *Parhyale* (Kao *et al*, unpublished) as well as another transcriptome incorporating leg tissue mRNA (Konstantinides, Semon and Averof, unpublished). Both of these resources were also used to obtain further details of *Parhyale*'s molecular clock.

5.2 Methods

5.2.1 The *Parhyale hawaiensis* head transcriptome

5.2.1.1 Sample collection and RNA extraction

A large sample of *Parhyale* were taken from the main culture and maintained in the same conditions with the exception of sparse crushed coral to ease collection of individuals. At eight time-points across 24 hours, beginning at 9am (lights on) and every three hours thereafter, a total of 20 adult males were collected per time-point; collections during the dark phase were undertaken using a head lamp providing dim red illumination. Individuals were caught using a plastic pipette, rinsed in DEPC-treated, sterilised distilled water, patted dry with a paper towel and then immediately frozen on a metal plate atop dry ice. The head was dissected in this frozen state using sterile scalpel blades and forceps, and the heads and bodies stored separately at -80°C until used for RNA extraction. All tools were treated with RNaseZAP (Sigma-Aldrich) between dissections, and the metal plate between time-point collections.

Total RNA was extracted from five heads per time-point using a micropestle to homogenise the tissue in TRIzol reagent (Invitrogen), cleaned using the PureLink Micro Kit (ThermoFisher Scientific) with on-column DNase treatment to remove genomic DNA contamination and eluted with RNase free water. RNA extractions were quality assessed using Nanodrop and Agilent 2100 Bioanalyzer. Samples failing to meet a minimum quality of > 50 ng/µL concentration; 260/280 ratio 1.8 − 2.2; 260/230 ratio 2.0 - 2.4 were discarded and the process repeated for that time-point with another five heads.

5.2.1.2 Illumina sequencing

One sample per time-point was delivered to Glasgow Polyomics (University of Glasgow) for library construction and 75 base pair paired-end strand-specific sequencing using the Illumina NextSeq 500 platform. The samples were subject to polyA mRNA enrichment prior to library construction with the TruSeq Stranded mRNA Library Prep Kit (Illumina). The raw data was subject to gentle low quality read removal and Q10 adaptor trimming by Glasgow Polyomics and subsequently retrieved via FTP.

Prior to assembly the reads from all time-points were combined into two files, one each for the forward and reverse reads of each pair, and FastQC 0.11.2 was then used to assess read quality. Cutadapt 1.9.1 was used to remove remaining traces of the TruSeq Index and Universal adapters using the sequence "GATCGGAAGAGC" common to both adapter types as criteria. Reads were subsequently filtered using Trimmomatic 0.32 (Bolger *et al.*, 2014) to remove reads with an average quality of less than 20 and/or shorter than 72 bases.

5.2.1.3 Assembly, quality assessment and annotation

A multi-assembler, multi-k-mer approach was adopted broadly as described at further length in 3.2.1.2, with some differences: the minimum and maximum k-mer used for Trans-ABySS was 19 and 71 respectively, the strand-specific option was used for each assembler and SOAPdenovo-Trans, lacking a strand-specific option, was not used at all. As outlined in sections 3.2.2.2 and 3.2.2.3 Transdecoder, CDHIT and Transrate were again employed in producing the final assemblies (total, coding and peptide), and BLAST and GO term annotation was performed against the arthropod UniProt Knowledgebase dataset and using BLAST2GO.

5.2.2 Searching for genes

5.2.2.1 Mining the transcriptomes

For the core circadian genes, the *Euphausia superba* orthologs were used as input queries against the total assembly using the tblastn function of BLAST+ with an E-value cutoff of $1.0e^{-3}$. Regulatory and output components were mined using both the *Euphausia superba* orthologs identified in Chapter Three and the input queries initially used to discover them as covered in section 3.2.3.2, again with an E-value cutoff of $1.0e^{-3}$.

Access was granted to a leg and embryo tissue transcriptome assembled using Trinity (Konstantinides, Semon and Averof, unpublished). Referred to hereafter as the KSA transcriptome, this was mined for core circadian genes using the same approach as above. The KSA transcriptome was also used for independent verification of the regulatory contigs identified in the head transcriptome.

5.2.2.2 Mining the draft genome

The draft genome is available at http://www.ncbi.nlm.nih.gov/genome/15533, accession GCA_001587735.1. It was mined for core circadian sequences using the *Euphausia superba, Eurydice pulchra, Drosophila melanogaster and Mus musculus* orthologs, and the *Parhyale* sequences identified from the transcriptome searches. Sequences showing a high degree of identity to the input queries were retrieved from the genome and translated, and any flanking bases that upon translation appeared to be in the same reading frame were included in subsequent steps. Sequences were subject to reverse-BLAST analysis against the NCBI NR database and the two *Parhyale* transcriptomes (the latter in the hope that where it was apparently lacking, evidence of expression might be obtained). Putative peptides were constructed from stretches of amino acids that obtained a relevant BLAST hit from NCBI NR and/or showed strong orthology to the initial search query.

5.2.3 Analysis of *Parhyale* orthologs

5.2.3.1 Confirmation and domain analysis

Each contig believed to represent a *Parhyale* circadian gene was subject to reverse-BLAST confirmation and core genes further characterised with SMART domain analysis as described in section 3.2.3.2. For each identified gene, peptide translations of the contigs derived from the head transcriptome, KSA transcriptome and genome were compared to provide independent verification of the assembled sequences.

5.2.3.2 Transcript abundance and expression analysis

Gene expression at each time-point was estimated using the expression matrix plug-in included with Trinity. The RSEM alignment method was used, aligning quality-trimmed reads against the 21 *k*-mer Bridger assembly, selected for the completeness of particular contigs of interest and Transrate score. For comparison across time-points, a TMM-normalised (Trimmed Mean of M values; Robinson and Oshlack (2010)) gene expression matrix was generated.

5.3   Results

5.3.1   Transcriptome assembly and annotation

5.3.1.1 Sequencing data and QC

A total of 192,807,920 paired end reads (385,615,840 total individuals reads) ranging from 32 – 75 base pairs (bp) in length was generated by the Next-Seq platform, approximately 24 million paired reads per time-point. After processing with Cutadapt 1.9.1 and Trimmomatic, 178,028,148 reads of length ≥ 72 bp remained with an average Phred quality of 20 or greater and a GC percentage of 47.

5.3.1.2 Assembly assessment

The **total assembly** of the head transcriptome consisted of 635,932 contigs, 97% of them rated as good quality, ranging from 200 to 35,494 bp in length, with an N50 of 2,063, 97% of fragments mapped and an optimised Transrate score of 0.5368. From this Transdecoder identified 94,050 **peptide sequences**, deriving from 86,921 contigs that make up the **coding assembly**.

5.3.1.3 Annotation

 Of the 86,921 contigs in the coding assembly, 51,977 returned a successful blast hit from the arthropod dataset. Of these, 37,004 were subsequently mapped with at least one GO term.

5.3.2   Transcriptome output

The results of mining the head and KSA transcriptomes are shown in Table 5.1. Complete sequences for *Parhyale hawaiensis bmal1* and *cry2* were discovered, while a fragment of *period* was also identified. There was no trace in either transcriptome of *clock* or *cry1*, and while *timeout* was identified, *timeless* was not.

Table 5.1: Output contigs encoding for putative core circadian peptides, from head and KSA transcriptomes.

Gene	Contig	Head transcriptome Contig Length	Peptide length	Type	Contig	KSA transcriptome Contig Length	Peptide length	Type
Phbmal1	PH.k21.comp127566_seq0	887	295	Internal fragment	c60121_g1_i1	2935	617	Complete
	PH.k19.S23178339	545	102	3' fragment				
	PH.k21.comp110465_seq0	326	55	5' fragment				
Phcry2	PH.k21.R17911710	3301	931	Complete	c33074_g1_i1	1450	482	Internal fragment
					c33074_g2_i1	614	160	5' fragment
					c179367_g1_i1	564	187	Internal fragment
					c87466_g1_i1	326	75	3' fragment
Phperiod	PH.k25.comp125155_seq0	280	91	3' fragment	c121112_g1_i1	256	85	Internal fragment
Phtimeout	PH.k21.comp121306_seq0	643	214	3' fragment	c47374_g1_i1	6544	1674	Complete
	PH.k21.comp62774_seq0	836	278	Internal fragment				
	PH.c77630_g1_i1	438	146	Internal fragment				
	PH.k31.comp64535_seq0	1478	360	Internal fragment				
	PH.k21.comp110010_seq0	228	76	Internal fragment				
	PH.k29.comp84379_seq0	506	168	Internal fragment				

5.3.2.1 *Phbmal1*

Mining the head transcriptome produced a number of contigs that reverse-BLAST analysis confirmed as encoding fragments of a putative PhBMAL1, including contigs covering the start and stop codons. The KSA transcriptome, meanwhile, output a 2,935 bp contig encoding a complete 617 aa peptide sequence (Figure 5.1), reverse-BLAST analysis returning *bmal1a* in the crayfish *Pacifastacus leniusculus* as the top hit (accession AFV39705: E-value $1e^{-150}$, identity 53%, coverage 75%). Domain analysis identified a bHLH domain, two PAS domains and associated PAC, a BCTR at the C-terminus and putative nuclear localisation and nuclear export signals.

```
MYSTGGYSNTHAEYISECGSIASVASLSSDGIAMKKKIPGHGECHNEDDLECSKLARSSAEWNKRQNHSEIEKRR
RDKMNTYISELSRMIPQCRSRKLDKLSVLRMAVQHIKMLRGSLNSYTEGQYKPGFVSDDEVQLLLKQECCESFLF
VVGCDRGKILFVSESVAHILQYTQQELLGSSWFDILHPKDLNKVKEQLSCGDMNRRERLVDAKTLLPVHQSPNSS
SSSGSSGNYPTLPADLTRLCPGSRRAFYARIRCPSVNKVQSDDGGGGGDSGSVCDESMTGDKRYLSIHFTGYLKS
WQGGRRASCSGGDDDHDSGDAACLVAIGRLHRPHADFPPLHFIAKLSAEAKYSYVDQRVSVVLGWLPQEILGASV
FELSHPSDHSTLSAAHRALLGKTCMAQSLHYRCRHKNGRWVQLTCKWTLFTNPWTNELEYIVASNSVLPSSPAPD
DALADASCRSAEPVISSPTVASVSLGGGSLDVRPASVTSCSYGGDAKAFTAGKDNLSSDALRTPHDGKGSGPTAR
DAADCRTGAHPQLSADTGETGDNSALPQQLDAPTRLPFHHHYNTRSESEASGVGETTSDSDEAAMAIIMSLLEAD
AGLGGPVDFSHLPWPLP
```
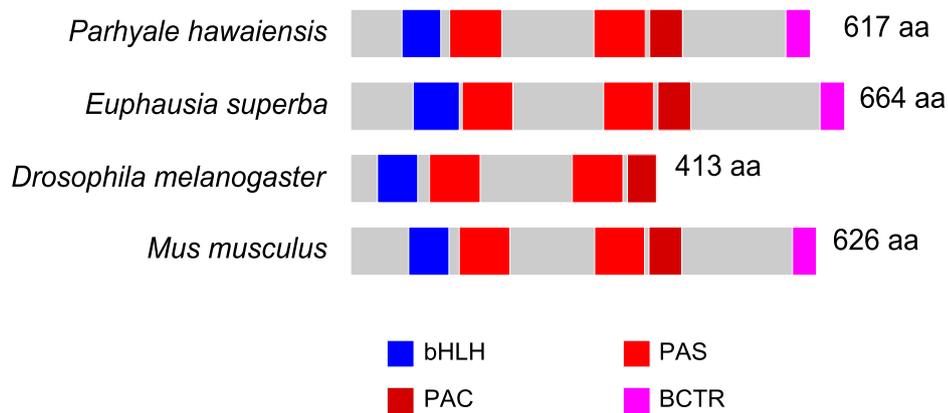

Figure 5.1: Top - Peptide sequence of putative *Parhyale hawaiensis* BMAL1. Turquoise = bHLH domain. Blue = nuclear localisation and export (white text) signal motifs. Red = PAS domain. Dark red/white text = PAC (motif C terminal to PAS) domain. Bottom - schematic comparisons of BMAL1/CYCLE peptides.

5.3.2.2 *Phcryptochrome2*

A 3301 bp contig was mined from the head transcriptome that encoded a putative complete PhCRY2 931 aa in length (Figure 5.2), while various fragmented containments were found in the KSA transcriptome. Reverse-BLAST analysis returned *cryptochrome 2* in the amphipod *Talitrus saltator* (accession AFV96168; E-value 0.0, 88% identity, 82% coverage). Domain analysis identified a DNA photolyase domain and a FAD-binding domain, both typical of cryptochrome proteins. Also of note was an extensive N-terminus. While it is not the first example of an extended CRY2 peptide - *Anopheles gambiae* also has a long extension at the C-terminus – it is the first identified N-terminus example of such magnitude. Alignment with the *T. saltator* ortholog suggested that 357 aa of this section was unique to *Parhyale*.

MGERNDPEVSISESSVHNVPAPVTGGQVMMVHPSRQSSEDMNVVASRSFISERFVPNRLTSDYPSMSSLPIMTNK
VLTGTVFMTDPSTMTQMTMVGDYKLESKALIPQSMLDVHPMYGDIKSIQGRAFIGEQMCSVQPMVTDALVLSGRT
MAPRLMPPKASPIKDDMMTGITMISGKPLSGRVMVGDARAVTTRIVTSDKATPVKTTTCEKASTNRPTDERTTPK
TSCATTVPKVHAAIEKHPPSKLSQIDRMTAARMVGDQPQQVRQEKKQTMAYKTGIKNYPIEKGNRTVTHEKQQLS
KNDSGEKPMVNERSSTTQPEVSRGEHFNFGCDSNAGPKLNSMTLERTKSPYQLAEPKMSPRKIASPGERSAQRTI
LGDNSVVQGDCTGVYGSSLPAEKKRCRVTPGKHVVHWFRRGLRLHDNPALRDSIINCETFRCIYILDPWFAGSSN
VGVNKWRFLLQCLEDLDNSLRKLNSRLFVVRGQPANALPQLFKEWNTTILSFEEDPEPFGRARDTSIIAIAQELG
IEVIVRTSHTLYKLDKIIEKKGGKPPLTYKTFQNILAMMDPPPAPVRPVVVDDLKFASTPLQPDHDDKYGVPNLE
HLGFETDNLPPAVWKGGETEALSRLKHHLERKAWVASFGRPKMTPQSLFACPTGLSPYLRFGCLSARKFYTELNE
LYIKIKKVPAVSLHGHLLWREFFYTAATNNPKFDHMKGNPICVQIPWDKNPEALAKWAHGQTGFPWIDAIMMQL
RKEGWIHNVARHAVACFLTRGDLWVSWEEGMKVFDELLLDADWSVNAGSWMWLSCSSFFQQFFHCYCPVRYGRKA
DPNGDFIRAYLPVLKNFPTKYIHEPWKAPEAVQRTARCLIGQHYPLPIVDHATQSQCNIERMKQVYQQLAHYRAN
ATSRSCSDTKGCFKSSGSGRPLTGGRMVTTV
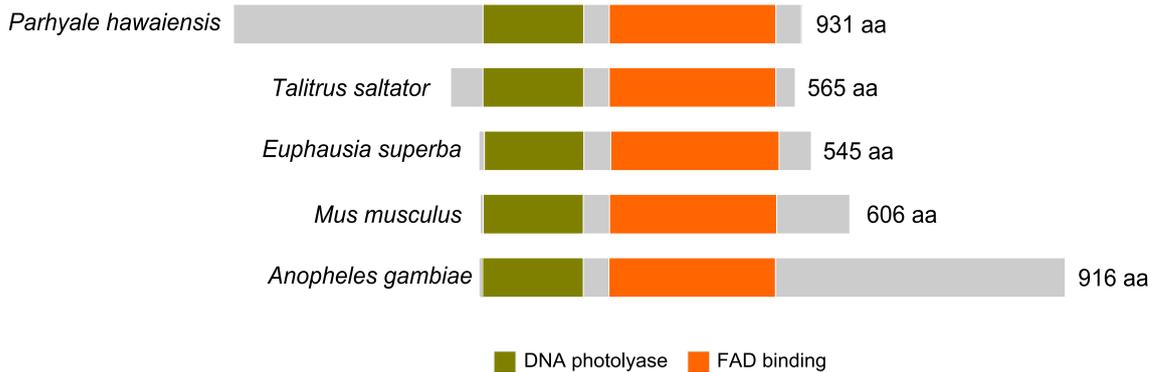
Figure 5.2: Top - peptide sequence of putative *Parhyale hawaiensis* CRY2 – red text highlights unique N-terminus. Olive green = DNA photolyase domain. Orange = FAD-binding domain. Bottom – schematic comparison with other CRY2 peptides, aligned at DNA photolyase domain.

5.3.2.3 *Phperiod*

Two small, overlapping fragments were obtained, one each from the head and KSA transcriptome. A construct of the two encodes a putative fragment of PhPERIOD, 131 aa in length (Figure 5.3). Reverse-BLAST analysis returned *period-like protein* in the lobster *Nephrops norvegicus* (accession ALC74274: E-value $1e^{-31}$, 53% identity, 84% coverage). Domain analysis revealed a PERIOD C domain, truncated at the 5' end.

*VEVTPQLLYRYQLRTKEIVDVLKNDMDALRELSQPALVEDQLSSLYQELEIDGEQLQLDEGITSSSGEEMVDAS
TKASSGNRRLEKIRSTRYFNKQAIIHDVEAAIPPPEMHVNHRYSVASARPVSIREIQ

Figure 5.3: Peptide sequence of putative *Parhyale hawaiensis* PERIOD. Green = PERIOD C domain.

5.3.2.4 *Phtimeout*

While no evidence was found of a 'true' *timeless* in either transcriptome, a 6,544 bp contig encoding a putative PhTIMEOUT, 1,674 aa in length, was identified from the KSA

transcriptome (Figure 5.4), while various containments were identified in the head transcriptome. Reverse-BLAST analysis returned *Timeless-like protein* in the termite *Zootermopsis nevadensis* (accession KDR17447: E-value 0.0, identity 49%, coverage 59%), but also - separately at the 3' end - a strong hit for *ubiquitin specific petidase 45* in *Mus musculus* (accession EDL05553: E-value $8e^{-36}$, identity 52%, coverage 10%). Domain analysis revealed a TIMELESS domain and a partial TIMELESS C domain, and an ubiquitin carboxyl-terminal hydrolase (UCH) domain at the C-terminus, which overlapped in part the TIMELESS C domain. None of the contained fragments obtained from searching the head transcriptome included this domain, while searching the head transcriptome using only the UCH domain sequence identified a contig that appeared to be a larger fragment of UCH 16, the top BLAST hit being that peptide in the copepod *Paracyclopina nana* (accession AII16561, E-value $2e^{-55}$). A fragment of PhTIMEOUT obtained from the head transcriptome, furthermore, deviates from KSA sequence at the point of a poly-aspartate sequence contained within the overlap of the TIMELESS and UCH domains. It therefore seems likely that the presence of the UCH domain is a misassembly.

```
MTAAIIEAELVAACSTLGYSDGKKYVRDPECLEVIRDLLRYLRRDDNYHQVRLALGETRVLQTDLVPLLREHHTD
FALLELLLRLLVNLTTPALLIFHQEIPEDKAGMQRYIKMQTQQQGFKEAFTEAAVWASIATVLSSRLQDGANRDP
ESENIIEMCLVLLRNVLSVPPSRQDSLRTSDDADVHDQVLWSLHLAGFPDLLLYLSSSTEESDLSLHTLEIISLM
LRQQDPQALATSALHRSAEEQKKDEAALVRARESEKARRQQTVRKHCSSRHSRFGGTYYVRNMKSISDRDIITHK
PLTDISAINFDENKRGKKVPKNRAPLPDSTTTRRSTLAIRLFLQEFCVEFLNGAYNNIMNIVKSNLDRARSQEHD
ESYYLWAMKFFMEFNRHHEFKVELVSETLSIQSVHYVQTNIETYHEMMTTERKKIPLWGRRMHNGLRAYQEILMT
LTAMDRCQDTAVRESSLALKSKLFYVVEYRELPLVLLLNYDPAKMSKGYLKDLVETTHVFLKLLEGMCKKTRQLM
VQLPEKKKNSAKKSKSTKKIIAPPTPEELEEKWGSLADELSAILQGEAGDLPVVVPFDALSELSEDEQKEKAMRK
VNALLRQSDLREAVALFRASREVWPEGDIFGAQDVDAPAEFACLREVFMAELNPVEQPENPEESEPEEDEEDELA
ASAVAESSAHISERELDFPAFVRRFAHIKILQSYSWLLKSYATNAPLTNHYIIKLFHRIAWDSKLPAVFFQSSLF
VTLHAAMTDPAKSSNEIIGQIAKFGKYIVRQFFQVAETNPKVFMELLFWKNYKEAVDIECGYDAPVATKAVKSLW
SEEEEDELTRLYEEFKEKVDPEGAKDLADHIMENLIRQDRSRRLVIKKLKDLGLISGLRELRNKPARVKGNQWSD
IELEELRVLFEEHKNDIDPMARILDFMVNPKPKHRVVEKLLELGIIQDKKEVRKKKIPKPKQQKAKKKGNNFGEQ
FLAANRGSDDEKSDVDTDEDSSSEDEEQSASTRNAPTPIPVVTPHLVSAALTKVRSLGFQESIQWLVEIMTEVAD
DREADNDFSAVPILAISEEQSKAIEDEDFQALMKIIGIQPPQSHEEMFWRVPEKLSVDNLRRRAQYLTQGLEGGL
PVSSSQDQSMSTSAPSAGSGDSGGAAAEVCSLVDGDQSQDDKENRNILNVGSVLNESSDELSDEQLPVSSGRDNS
RASVVMPKRPLEDHSANFDALLSSQPPKKTKKRRIVVDNSDDDDDDDDDDEDEEDSDEKNGESDAASLAKKKENS
DADEEDNAESELPRERVTRQPSIIVVEQMEHGDSGCPNSASDGRYLSTSASCDHLNYNSASDTSERYFTADGNGI
DTGDDLDTPKRRSLSKECLLKASTDEELARRLCELKLHRKSSSKESLHRLNFKRPSAETLQTFRSCERLTEEDDY
GLTMEANPKKFKQEWLARSLTTLSPKYQAKAQECSVMSCLAKFTAPELLASNNKIICQNCTKLRNISGNNPEKDG
PVRSPARKQLMVVCPPAVLTLQLKRFHHDGVHLAKANRFVHFPLVLDLAPFTSTIAMSPGTRLLYGLYGIVQHTG
RLHNGHYTAFVRSRPITANRPPPTAYLTQHPLDSASLETDPPPRLPPSKTATTVPAVEELEAEIRKDQWYHVSDA
HVTHVTIDRVLKAQAYLLFYERIV
```

Figure 5.4: Peptide sequence of putative *Parhyale hawaiensis* TIMEOUT. Black = TIMELESS domain. Purple = TIMELESS C domain. Red = UCH domain. Blue = overlap of TIMELESS C and UCH – note the poly-D sequence within this overlap (see text for discussion).

5.3.2.5 Regulatory genes

Table 5.2 shows the output from each transcriptome and the peptides derived from whichever version was deemed the most complete in each case. Candidates were identified for every *Parhyale* regulatory and output protein with the exception of JETLAG and either form of PDH. Of those successfully identified, all but three were present and in agreement in both the head and KSA transcriptome. Table 5.3 shows the results of blastp analysis against the NCBI non-redundant database and Flybase.

Table 5.2: Transcriptome mining: clock-related query protein details and *Parhyale hawaiensis* output contigs. Red highlight indicates contig rejected as ortholog after reverse-BLAST. Blue highlight indicates preferred contig where output of the transcriptomes varied. 'Single contig coverage' – percentage of contig identified in one genome contig. [a] Fragment.

Query Protein		*Parhyale hawaiensis* transcript/protein identifications									
		Head transcriptome output			KSA transcriptome output			Present in	Single contig	Protein	
Protein name	Query	Contig ID	E-value	Length	Contig ID	E-value	Length	genome	coverage	Name	Length
AANAT	AAM68307	PH.c136570_g2_i4	2.00E-16	2412	-	-	-	Yes	100%	PhNAT	297
CASEIN KINASE II α	EsCKIIα	PH.k29.comp2312_seq4	0	4884	c269302_g1_i1	0	2985	Yes	88%	PhCKIIα	352
CASEIN KINASE II β	EsCKIIβ	PH.k27.comp3841_seq0	5E-129	1477	c57024_g1_i1	2E-129	1510	Yes	94%	PhCKIIβ	222
CLOCKWORK ORANGE	EsCWO	PH.k21.comp80764_seq0	1E-35	642	c86247_g1_i1	1E-24	550	Yes	75%	PhCWO	212[a]
CTRIP	EsCTRIP	PH.k31.comp1615_seq0	0	12125	c49661_g3_i1	0	12479	Yes	85%	PhCTRIP	2996
DOUBLETIME	EsDBT	PH.k27.comp42178_seq0	0	1599	c23376_g1_i1	0	5983	Yes	86%	PhDBT	418
JETLAG	AAF52178	PH.k51.S3954173	1E-12	4230	c56926_g1_i1	5E-13	4620	-	-	-	-
LARK	EsLARK	PH.k31.comp1313_seq3	4E-84	1416	c56502_g1_i1	2E-82	2021	Yes	68%	PhLARK	303
NEJIRE	EsNEJ	PH.k25.comp26397_seq3	0	5844	c37596_g1_i1	0	5142	Yes	92%	PhNEJ	1777[a]
NEMO	EsNEMO	PH.c118708_g1_i2	0	1416	c13423_g1_i2	2E-176	3392	Yes	68%	PhNEMO	554[a]
PAR DOMAIN PROTEIN 1 ε	EsPDP1	PH.k21.comp5330_seq6	5E-08	3178	c47597_g1_i2	5E-33	3709	Yes	52%	PhPDP1	591
PIGMENT DISPERING HORMONE	EsPDH1	-	-	-	-	-	-	-	-	-	-
	EsPDH2	-	-	-	-	-	-	-	-	-	-
PDH RECEPTOR	EsPDHR	PH.k21.comp140546_seq0	3E-43	597	c276158_g1_i1	6E-50	415	Yes	98%	PhPDHR	187[a]
PROTEIN PHOSPHATASE 1	EsPP1	PH.c136613_g1_i3	0	2766	c14544_g2_i1	0	3349	Yes	98%	PhPP1A	329
PP2A - MICROTUBULE STAR	EsMTS	PH.k21.comp640_seq0	0	3538	c100372_g1_i1	0	3529	Yes	89%	PhMTS	309
PP2A - WIDERBORST	EsWBT	PH.k41.S6405731	0	2831	c177636_g1_i1	0	2881	Yes	51%	PhWBT	458
PP2A - TWINS	EsTWS	PH.k21.R18132040	0	2887	c52458_g1_i1	0	5227	Yes	72%	PhTWS	444
REV-ERBα/E75	EsE75	PH.k31.comp813_seq0	1E-99	4901	c35279_g2_i1	2E-100	5901	Yes	78%	PhE75	866
RORA	EsHR3	PH.k51.R3934289	2E-70	3025	c58020_g1_i1	2E-70	3557	Yes	62%	PhHR3	879
SHAGGY	EsSGG	PH.c140539_g2_i4	0	2874	c23335_g1_i1	0	5108	Yes	90%	PhSGG	426
SUPERNUMERARY LIMBS	EsSLIMB	PH.k25.comp29022_seq3	0	4791	c270655_g1_i1	0	2838	Yes	59%	PhSLIMB	642
TAKEOUT	EsTAKEOUT	PH.k21.comp2754_seq2	1E-50	4655	c57993_g1_i1	1E-51	3042	Yes	100%	PhTAKEOUT	272
VRILLE	EsVRI	PH.k21.comp5330_seq6	1E-52	3178	c270979_g1_i1	1E-53	2833	Yes	100%	PhVRI	480

Table 5.3: blastp analyses of putative *Parhyale hawaiensis* clock-associated proteins against *Drosophila melanogaster* protein database (Flybase) and NCBI non-redundant protein database. [a] Fragment.

Query	Top Flybase hit			Top NCBI nr hit			
	Flybase No.	Associated gene name	E-value	Accession	Name	Species	E-value
PhNAT	FBpp0291637	*CG13759-PD*	1E-24	NP_570009	CG13759, isoform A	*Drosophila melanogaster*	7E-08
PhCKIIα	FBpp0070041	*casein kinase IIa*	5E-160	EFN84867	Casein kinase II subunit alpha	*Harpegnathos saltator*	0
PhCKIIβ	FBpp0300428	*Casein kinase II ß*	3E-109	ACO10984	Casein kinase II subunit beta	*Caligus rogercresseyi*	0
PhCTRIP	FBpp0306924	*circadian trip*	4E-119	KYB27699	E3 ubiquitin-protein ligase TRIP12-like Protein	*Tribolium castaneum*	0
PhCWO[a]	FBpp0081723	*clockwork orange*	1E-28	KOC64192	Hairy/enhancer-of-split related with YRPW motif protein 1	*Habropoda laboriosa*	4E-50
PhDBT	FBpp0306615	*discs overgrown*	9E-154	AGV28719	Casein kinase 1 epsilon	*Eurydice pulchra*	0
PhE75	FBpp0312446	*Ecdysone-induced protein 75B*	7E-94	KYM84307	Ecdysone-inducible protein E75	*Atta colombica*	5.00E-97
PhLARK	FBpp0076555	*lark*	2E-58	BAN20619	arginine/serine-rich-splicing factor	*Riptortus pedestris*	2E-62
PhNEJ[a]	FBpp0305701	*nej*	0	KDR19833	CREB-binding protein	*Zootermopsis nevadensis*	0
PhNEMO[a]	FBpp0305425	*nemo*	3.67E-166	CAH65680	nemo-like protein	*Nilaparvata lugens*	0
PhPDP1	FBpp0289727	*PAR-domain protein 1*	4E-33	ABV22507	PAR domain protein 1	*Danaus plexippus*	2E-46
PhPDHR[a]	FBpp0309084	*Pigment-dispersing factor receptor*	2E-35	BAH85843	pigment dispersing hormone receptor	*Marsupenaeus japonicus*	2E-149
PhPP1	FBpp0306442	*Protein phosphatase 1a at 96A*	8E-179	KYQ51663	Serine/threonine-protein phosphatase alpha-1 isoform	*Trachymyrmex zeteki*	0
PhMTS	FBpp0310063	*microtubule star*	6E-174	KDR18186	PP2A catalytic subunit alpha	*Zootermopsis nevadensis*	0
PhWBT	FBpp0084575	*widerborst*	0	KYB26642	PP2A regulatory subunit epsilon	*Tribolium castaneum*	0
PhTAKEOUT	FBpp0307590	*CG10407*	2E-26	ACO14761	takeout precursor	*Caligus clemensi*	6.00E-08
PhTWS	FBpp0081671	*twins*	0	AFK24473	PP2A regulatory subunit B	*Scylla paramamosain*	0
PhHR3	FBpp0297438	*Hormone receptor-like in 46*	3.74E-70	EFN70388	Probable nuclear hormone receptor HR3	*Camponotus floridanus*	6.00E-111
PhSGG	FBpp0070450	*shaggy*	0	ALK82316	glycogen synthase kinase 3 beta	*Macrobrachium nipponense*	0
PhSLIMB	FBpp0306059	*supernumerary limbs*	0	KDR19729	F-box/WD repeat-containing protein 1A	*Zootermopsis nevadensis*	0
PhVRI	FBpp0312171	*vrille*	2E-41	CAX37108	Vrille	*Acyrthosiphon pisum*	1E-63

5.3.3   Genome output

5.3.3.1 *Phbmal1*

The contig identified from the transcriptome was supported by genome searches. Genome contig phaw_30.0000271 (accession LQNS01000271) showed sequential fragmented hits from beginning to end of the transcriptome contig, with no large omissions, in the range covered by position 39,836 – 71,014. A further small section was found on phaw_30.0097676 (LQNS01097676).

5.3.3.2 *Phclock*

Evidence of *Phclock* was found in the genome on contigs phaw_30.0005030 and phaw_30.0012642 (LQNS01005030 and LQNS01012642). Fragmented hits within the range covered by position 204,755 – 219,765 (phaw_30.0005030) and 55,926 – 56,078 (phaw_30.0012642) showed a high degree of identity with the query peptides. The peptide construct is shown in Figure 5.5: reverse-BLAST analysis identified the partial CLOCK of *Pacifastacus leniusculus* as the top hit (accession AFV39704: E-value $1e^{-88}$, identity 72%, coverage 96%), while the ortholog of *Macrobrachium rosenbergii* was the best hit from a complete sequence (accession AAX44045: E-value $1e^{-86}$, identity 73%, coverage 94%). Domain analysis identified a bHLH domain and a PAS and associated PAC domain. Study of the specific amino acid sequence confirmed that the PAS domain is likely to be the second of the two typically seen in CLOCK proteins. No further sequence data was identified in the genome, and searching both *Parhyale* transcriptomes using this genome derived sequence found no further incidence of *Phclock*, supporting the earlier finding of no expression data for this gene.
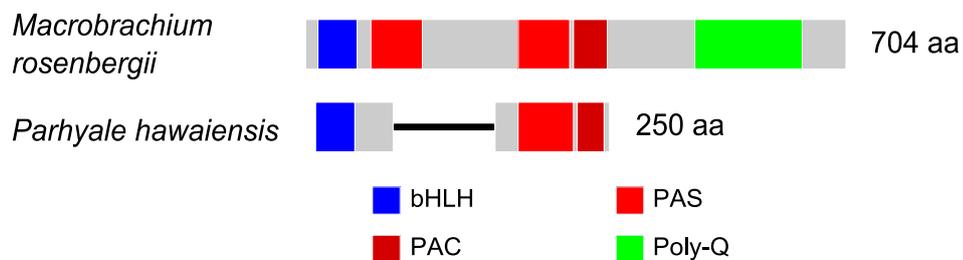
Figure 5.5: Fragmented putative PhCLOCK peptide constructed from mined genome sequences and aligned with the CLOCK of *Macrobrachium rosenbergii* (best BLAST hit with a complete sequence).

5.3.3.3 *Phperiod*

Further *Phperiod* sequence data was obtained from the genome, on contig phaw_30.0001982 (LQNS01001982). A series of fragmented hits from within the range covered by position 735,402 – 764,473 showed a high degree of identity with the query peptides. The peptide construct is shown in Figure 5.6: domain analysis identified two PAS domains and a PAC domain, as well as the PERIOD C domain already seen in the transcriptome contig. BLAST search showed the sequence to be most similar to *period* in *Eurydice pulchra* (accession AGV28714; E-value 4e^{-74}, identity 59%, coverage 80%).
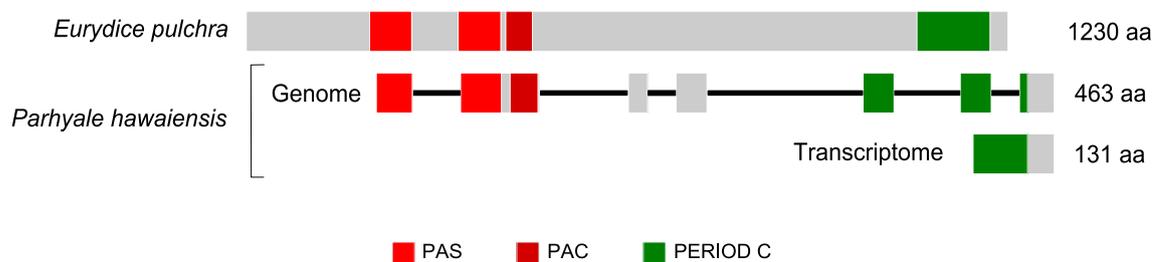
Figure 5.6: Fragmented putative PhPERIOD peptide constructed from mined genome sequences and shown in comparison to *Eurydice pulchra* PERIOD and the assembled transcriptome sequence. Note that the distance between genomic fragments are estimated based on alignments with *Eurydice pulchra* PERIOD, within the exception of the final gap where alignment with the transcriptome sequence was possible.

Searching both *Parhyale* transcriptomes using this genome derived sequence found no further incidence of *Phperiod* – the small fragment already identified is the only representation of the gene in the expression data.

5.3.3.4 *Phtimeless*

No trace of a 'true' *timeless* was found in the genome, with all resulting hits proving to be part of the previously identified *Phtimeout*. In further support of the idea that the UCH domain seen in PhTIMEOUT is due to a misassembly, searching the genome with that peptide sequence returned coverage on a single contig, phaw_30.0002198 (LQNS01002198), only up to the beginning of that domain, with the UCH sequence found elsewhere.

5.3.3.5 *Phcryptochrome*

No *Drosophila*-like cryptochrome was identified in the genome. The contig identified as *Phcry2* from the transcriptome was strongly supported by genome searches, identified in full on a single contig, phaw_30.0004738 (LQNS01004738), in the range covered by positions 265,571 – 278,322. Notably, the range 276,943 – 278,322 encodes an uninterrupted peptide sequence that encompasses the large N-terminus and the beginning of the DNA photolyase domain. Taken with one of the internal fragments mined from the KSA transcriptome this feature is supported by two sources of independent verification (Figure 5.7).
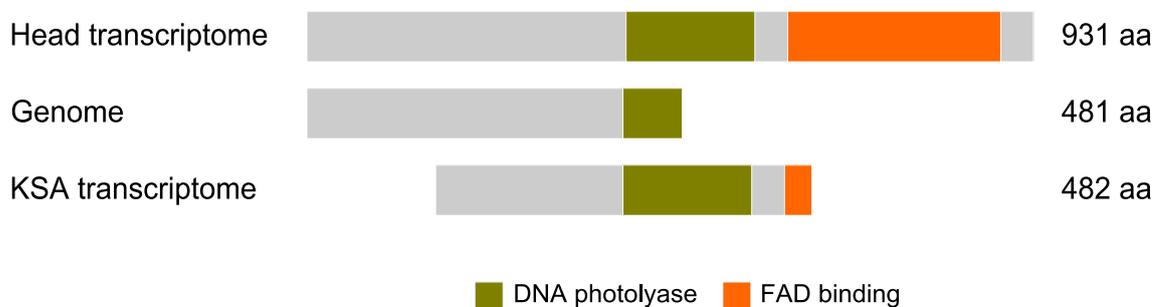
Figure 5.7: Schematics of putative PhCRY2 peptides from three independent sources, aligned. Top - the head transcriptome. Middle – the first uninterrupted sequence found in the genome. Bottom – an internal fragment identified in the KSA transcriptome. Other genome and KSA fragments omitted for clarity.

5.3.3.6 Regulatory genes

Evidence was obtained supporting all regulatory and output genes identified through the transcriptomic approach, the majority of them with a high degree of coverage on a single genomic contig. The genome was also searched for both forms of PDH using orthologs from a number of insects and crustaceans as queries, without success.

5.3.4   Gene expression

5.3.4.1 Core genes

Expression of the identified core circadian genes *Phbmal1*, *Phperiod* and *Phcry2* was low across all time-points. The former two in particular showed no evidence of

expression (Figure 5.8) at certain time-points, with *Phperiod* registering in only the 9 PM and 3 AM samples.

5.3.4.2 Regulatory and output genes

Taking in consideration only those genes considered highly likely to contribute to regulation of the *Parhyale* clock – that is, excluding output genes, the more speculative genes mined on the basis of mammalian clocks, and *timeout*, whose contribution to the clock remains controversial – it is possible to discern a pattern in the expression data for this set of genes in which expression peaks at 12 - 3 PM and more pointedly at 3 AM (Figure 5.9). This pattern is also evident in the expression data of *Phcry2* (included in the above figure for comparison)*,* the output gene *Phpdhr* and the putative arylalkylamine N-transferase *Phaanat.* A *Parhyale ribosomal L32* gene identified from the head transcriptome, which might be expected to show stable expression over 24 hours, is included for reference. A repeated measures comparison using Friedman's test showed that normalised expression of these regulatory genes varied significantly depending on time-point (χ^2(7) = 56.095, p < 0.001). The same test performed on an equal number of randomly selected genes showed no significant effect of time-point (χ^2(7) = 9.855, p = 0.197) and this was true for two further repeated tests of randomly selected genes (χ^2(7) = 11.626, p = 0.114; χ^2(7) = 9.316, p = 0.231). Note that this test still returned a significant result when returning the previously excluded genes to the dataset (χ^2(7) = 54.137, p < 0.001).

The actual TMM-normalised values across time-points for the regulatory and output genes are shown in Table 5.4. The majority of clock-related genes appear to show relatively low levels of expression.
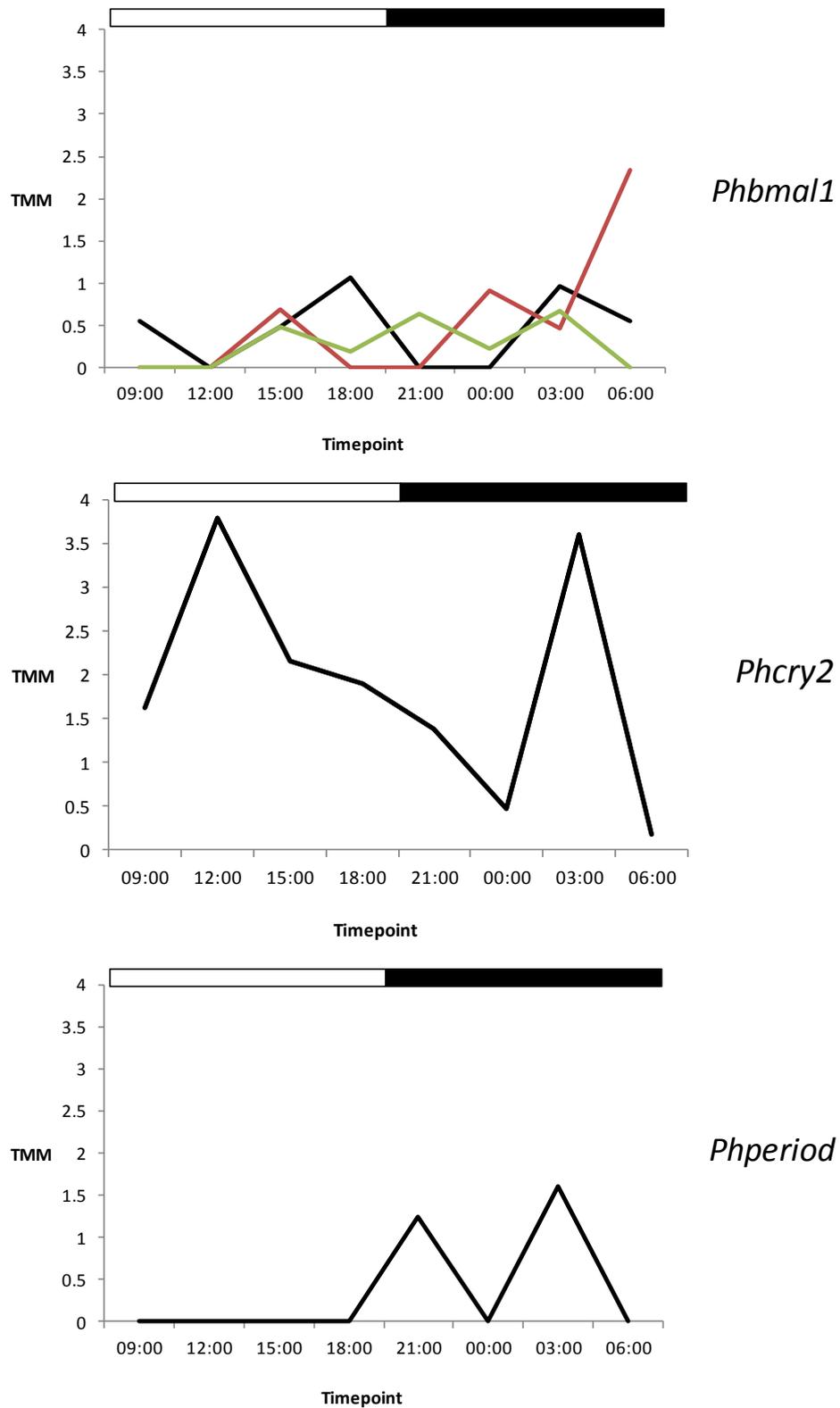
Figure 5.8: TMM-normalised expression values for the core circadian genes identified in the *Parhyale* head transcriptome. *Phbmal1* was fragmented in the transcriptome – here TMM for all three fragments is shown. White and black bars show light and dark phase over 24 hours.
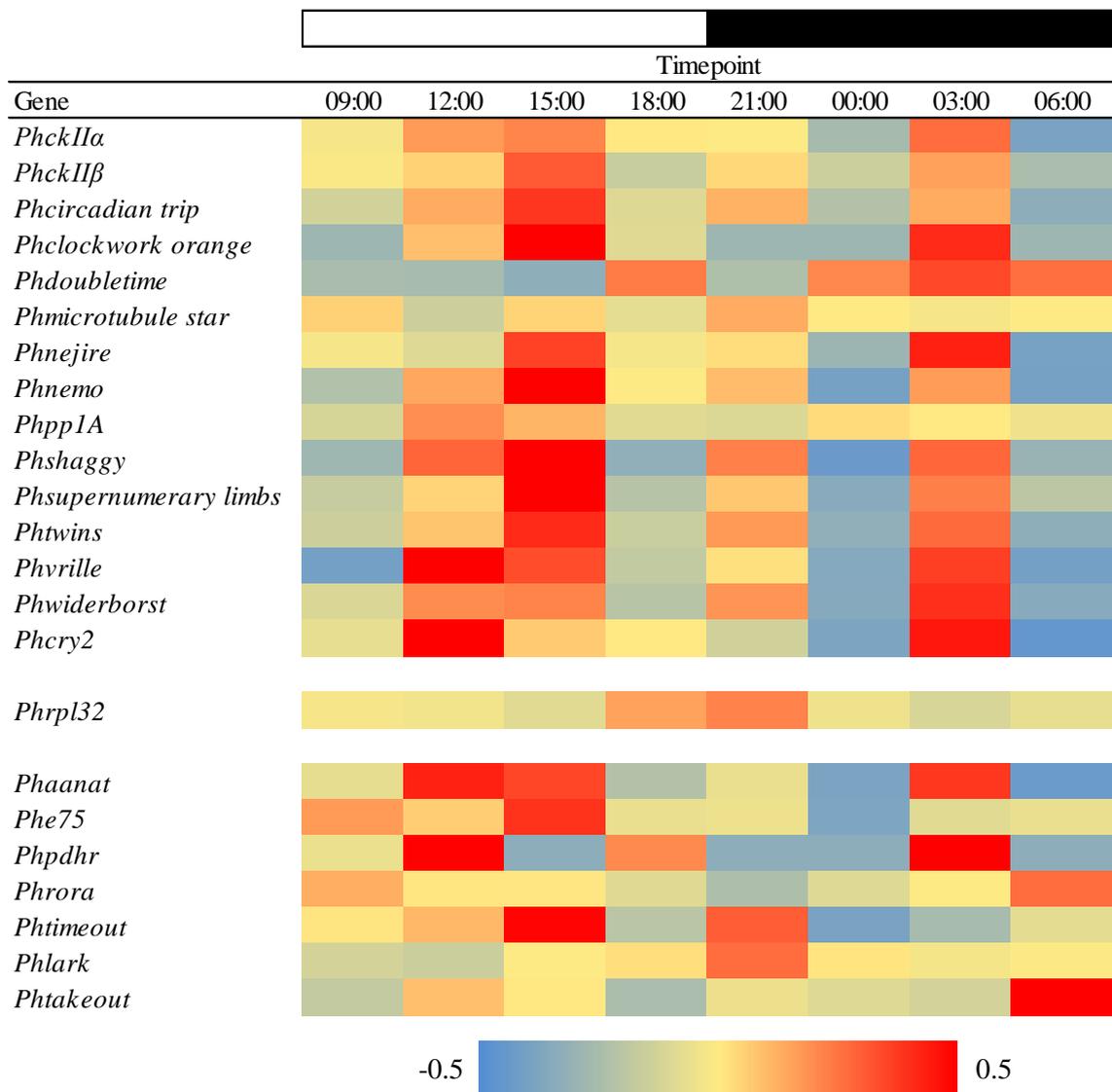
Figure 5.9: For each gene data was normalised to the maximum expression value (so maximum = 1), and the mean of normalised values calculated. Each spot is then colour-coded according to the deviation from the mean at each time-point. First grouping shows genes considered highly likely to contribute to regulation of the clock. Second grouping shows putative output genes and those identified on the basis of orthology to mammalian circadian genes. *Phrpl32* is a putative ribosomal L32 gene.

Table 5.4: TMM-normalised expression values of putative regulatory and output circadian components of *Parhyale hawaiensis* across 24 hours. Grey cells highlight zero expression. White and black bars show light and dark phase.

				Timepoint				
Gene	09:00	12:00	15:00	18:00	21:00	00:00	03:00	06:00
Phaanat	3.15	6.31	5.82	2.21	3.21	1.21	6.03	0.90
PhckIIα	4.49	5.76	6.04	4.71	4.62	3.00	6.39	2.18
PhckIIβ	26.16	28.73	38.58	20.48	28.24	20.83	32.77	17.30
Phcircadian trip	7.03	10.93	14.56	7.53	10.74	5.72	10.94	4.15
Phclockwork orange	0.00	0.55	1.44	0.28	0.00	0.00	1.00	0.00
Phdoubletime	2.15	2.10	1.72	4.68	2.21	4.55	5.26	4.82
Phe75	86.26	74.26	110.98	60.43	61.54	25.63	57.52	60.48
Phlark	21.21	20.27	25.68	26.72	35.26	26.24	24.65	25.44
Phmicrotubule star	118.69	92.43	117.83	101.80	129.00	112.20	108.30	111.24
Phnejire	1.60	1.40	2.75	1.60	1.78	0.83	2.96	0.52
Phnemo	0.17	0.54	0.99	0.39	0.49	0.00	0.56	0.00
Phpdhr	0.15	0.46	0.00	0.30	0.00	0.00	0.55	0.00
Phpp1A	21.88	31.97	29.31	22.78	22.31	26.77	25.82	24.25
Phrora	4.37	3.77	3.76	3.24	2.51	3.22	3.70	5.06
Phshaggy	3.42	13.31	17.56	2.63	12.29	0.73	13.28	3.09
Phsupernumerary limbs	0.45	0.79	1.60	0.38	0.84	0.16	1.08	0.40
Phtakeout	10.51	22.65	18.80	7.42	15.93	14.09	12.53	45.11
Phtimeout	0.99	1.17	1.90	0.58	1.54	0.23	0.48	0.82
Phtwins	0.80	1.22	1.81	0.79	1.39	0.49	1.56	0.48
Phvrille	0.00	2.70	1.99	0.60	1.15	0.13	2.07	0.00
Phwiderborst	2.52	4.08	4.17	2.01	3.99	1.26	5.09	1.28

5.4 Discussion

5.4.1 Summary of results

The full sequences encoding two core circadian proteins, PhBMAL1 and PhCRY2, have been identified in *Parhyale hawaiensis,* the former from an existing embryo and leg tissue transcriptome and the latter from a head transcriptome, the assembly of which is recounted here. While the PhBMAL1 is a typical example of a BMAL1 protein, PhCRY2 possesses an extensive N-terminus that is unique to this species – both sequences are well-supported with evidence in two independent transcriptome assemblies and the recently completed genome.

Regarding other core genes, only a small fragment of *Phperiod* has been identified in the transcriptomes, while further sequence data has been obtained from the genome, albeit still fragmented. There is no sign of *Phclock* in the transcriptome, while genome mining has produced only a small amount of fragmented evidence. There is no evidence in either transcriptome or in the genome of a *Drosophila*-like *cryptochrome* or a 'true' *timeless*.

Putative orthologs of all regulatory and output genes searched were obtained from the transcriptomes with the exception of *jetlag* and *pigment dispersing hormone*, neither of which appeared to be present in the genome either.

Expression analysis suggests low expression levels of many of the circadian genes including the core elements, with *Phbmal1* and *Phperiod* in particular failing to show any expression at certain time-point. Looking at the expression of regulatory genes as a group a bimodal pattern can be discerned, peaking at 12 – 3 PM and 3 AM.

5.4.2 Core genes

5.4.2.1 Those that are present...

PhBMAL1 again confirms the near-ubiquity of the BCTR and is an unexceptional example of the circadian protein except for the notably low levels of underlying gene expression, with some time-points showing no evidence of expression at all. This phenomenon is seen even more starkly with *Phperiod*, which is detected only in two dark phase time-points (Figure 5.8) and is in such low abundance that only a small

fragment of the C-terminus stood as evidence in the head transcriptome assembly. Even the completely assembled *Phcry2* shows very low levels of expression at certain time-points, and it is also notable that a full coding sequence for PhBMAL1 was obtained from the KSA transcriptome, derived from embryonic and leg tissue, and seen only in fragmented form in the head transcriptome, while the opposite is true for PhCRY2. This will be addressed further in 5.4.4.

PhCRY2 appears to possess a unique, extended N-terminus (Figure 5.2), which was independently verified in a separate transcriptome and in the genome. Various analyses detected no orthology to any known peptide, nor any putative domains, motifs or signals. It may not have a purpose, and the activity of the peptide may proceed unhindered – *Anopheles gambiae* CRY2, despite its long C-terminus, is able to function as a potent repressor of CLK:CYC/BMAL1 heterodimers (Yuan *et al.*, 2007). Nevertheless it is an intriguing feature warranting further investigation.

5.4.2.2 ...and those that are not

It is hard to draw a firm conclusion regarding *Phclock*. With no evidence found in any transcriptome and minimal fragmentary evidence in the genome, the temptation is to conclude that the gene has been lost and the fragments seen in the genome are pseudogene remnants. The loss of this component would not necessarily be the undoing of a functional molecular clock in *Parhyale,* as CYC/BMAL1 proteins have proven to be rather promiscuous in forming transcriptionally active dimers with other bHLH-PAS proteins (Hogenesch *et al.*, 1998) and more specifically such atypical heterodimers have been found to substitute for CLOCK in circadian scenarios, such as NPAS2:BMAL1 in mice (DeBruyne *et al.*, 2007) and methoprene-tolerant (MET):CYC in the mosquito *Aedes aegypti* (Shin *et al.*, 2012). An ortholog of MET, as well as other bHLH-PAS genes such as *hypoxia inducible factor*, have been identified in the head transcriptome and could in theory substitute for CLOCK in *Parhyale*. This path of enquiry is well worth pursuing.

All of that said, there *is* evidence of *Phclock* in the genome just as there is for *Phperiod*, which is found in the transcriptome at such low levels that only a small fragment was assembled from reads obtained from two time-points out of eight. It would not take

much lower expression levels of *Phclock* for it to appear entirely absent. It is also noticeable that the identified fragments both cover conserved domains – the rest of the sequence may be too derived to identify easily through BLAST homology. Searching the genome with *Eurydice pulchra* BMAL1 and CLOCK produces 52% coverage for the former and just 16% for the latter, but EpBMAL1 is only 64% the length of EpCLOCK despite possessing similar sized bHLH and PAS domains and a highly conserved BCTR and so proportionally, a greater amount of its sequence consists of conserved, and thus more easily identifiable, domains. The genome itself may also require further development before some genes can be successfully identified in their entirety – many of the genomic fragments identified using transcriptomic or orthologous sequences such as PhBMAL1 or *Eurydice pulchra* PER were short and separated by thousands, sometimes tens of thousands, of base pairs. *Phperiod*, which has evidence of expression in the transcriptome, was still identified only in fragmentary form and mostly at the sites of conserved domains.

Regarding *timeless* and *cry1*, the complete absence of both transcriptomic and genomic evidence leads to the conclusion that they have been lost. In this aspect *Parhyale*'s clock reflects that of the hymenopterans (Tomioka and Matsumoto, 2015) and mammals (Reppert and Weaver, 2000), which lack both genes.

5.4.3   Regulatory and output genes

As with the output from the *Euphausia superba* transcriptome database, the identification of so many putative regulatory and clock-controlled components leaves us with an abundance of potential research paths, particularly in a tractable model organism such as *Parhyale* – for example it will be possible to assess the effect of RNAi knockdown of these genes. Also worthy of further investigation is the apparent bimodal pattern of expression many regulatory genes, and *Phcry2*, seem to exhibit. The major aim in producing the head transcriptome was gene identification rather than expression analysis and the decision to use time-point samples was based on the fact that such a design cost little more than a one-sample run producing the same number of reads, while offering a greater degree of information. While no robust conclusion can be drawn about the expression of any individual gene without replicates, particularly at such low expression levels, it is striking that a significant

pattern is seen across time in a suite of genes believed to regulate the circadian clock, which is not seen in randomly selected groups of genes.

Interestingly, *cry2* in the mosquito *Aedes aegypti* shows a bimodal expression pattern under 12:12 LD, with peaks around 3 AM and 5 PM (Gentile *et al.*, 2009). As only one replicate is available for each time-point, and the sampling covers only one 24 hour period, no decisive test of the significance of this pattern can be performed with the current data. However RT-PCR analysis could lend further supporting evidence, or further sampling across multiple days would allow the identification and confirmation of a repeating pattern.

A notable absence is that of the *Pdh* gene, the *Drosophila* ortholog *pdf* being closely tied to the output of the clock and *Pdh* itself linked to circadian light responses in *Procambarus clarkii* (Verde *et al.*, 2007). No trace can be found in either transcriptome or in the genome, both through BLAST search and via annotation of peptide and coding sequences, and despite the presence of its receptor *Pdhr* in the transcriptome it may have been lost.

5.4.4 Where is it?

5.4.4.1 Regarding the head transcriptome

The title of this section and indeed the subtitle of this whole chapter, "Where is it?", reflects not only the author's (sanitised) initial reaction upon surveying the distinctly bare results of searching the head transcriptome for core circadian genes but also (upon reaching acceptance) the less rhetorical issue of what the implications are if the data accurately reflect the reality. If clock genes are truly expressed at very low levels in the head, or in some cases not at all, then where is the *Parhyale* clock? Firstly, though – do the data indeed accurately reflect the reality?

The KSA transcriptome, assembled from reads deriving from embryonic and leg tissue, provided a complete coding sequence for PhBMAL1. Over 70% of this was identified in fragmentary form in the head transcriptome, but one might expect the head transcriptome to at least match the output of leg and embryo tissue with regards to circadian genes. The head transcriptome does not lack for raw data with nearly 180

million paired end reads (nearly 360 million in total) used for assembly. For comparison, the *Euphausia superba* transcriptome was assembled from 35 million read pairs. It appears to be well-assembled, judging by its high Transrate score and percentage of mapped fragments. The majority of genes searched for have full coding sequences and have been confirmed to be present in the same form in the independently assembled KSA transcriptome, and further searches for previously published Hox gene fragments (Martin *et al.*, 2015) also produced full coding transcripts, as did searches for opsins (thankfully – if you cannot find an opsin in a head transcriptome something has definitely gone awry). Furthermore, the failure to identify certain genes does not appear to be a failure of assembly. No evidence of *timeless* is found even in the genome, and the further fragments of *Phperiod* and small fragments of *Phclock* that are found there are conclusively not present in either the head or KSA transcriptome – they are simply not there, rather than poorly assembled or fragmented and thus hard to find.

An illustration of the head dissection that provided tissue for mRNA extraction is shown in Figure 5.10. This would have certainly included the brain, eyes, optic and antennal nerves, all of which are good candidate sites for circadian oscillator expression. It is possible that the suboesophageal ganglion was omitted or only partly taken in some dissections, but this is not typically implicated in crustacean circadian rhythms. It seems clear from the figure that the tissue taken would represent the best chance for obtaining gene sequence data for a system typically most abundant in the brain.
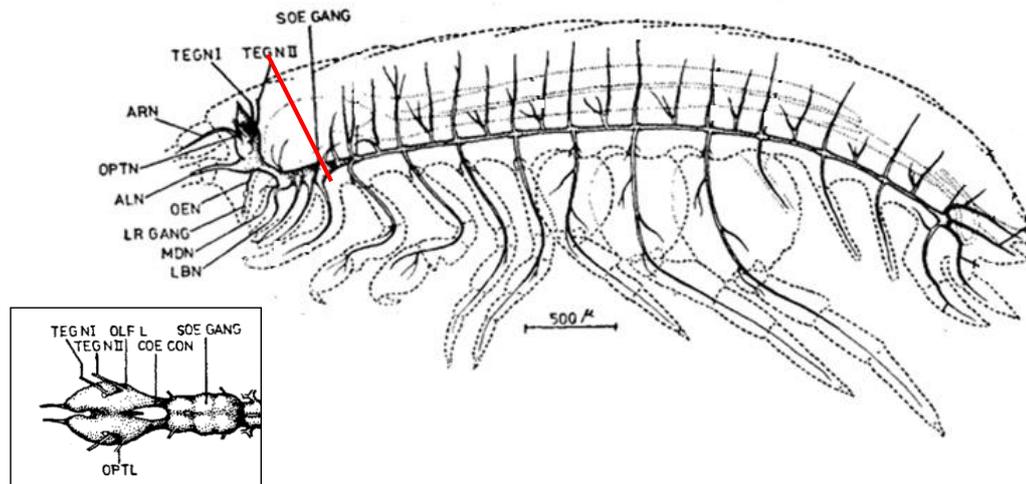
Figure 5.10: Nervous system of *Parhyale hawaiensis*. Main image – lateral view, red line indicates line of head dissection. Inset – dorsal view of brain and suboesophageal ganglion. Modified from Divakaran (1982) with relevant labelling retained. ALN – antennal nerve; ARN – antennular nerve; COE CON – circumoesophageal connective; LBN – labial nerve; LR GANG – labral ganglion; MDN – mandibular nerve; OEN – oesophageal nerve; OPTL – optic lobe; OPTN – optic nerve; SOE GANG – suboesophageal ganglion; TEGN – tegumental nerve.

Perhaps the use of the entire head is an issue. The most abundant transcript with an average TMM-normalised expression value of 39,080, more than double that of the next most abundant, represents a *rhodopsin* gene (top BLAST hit *rhodopsin* in the shrimp *Litopenaeus vannamei*, accession ABH00987, E-value 0.0). In typical RNA-seq the cDNA libraries are not normalised prior to sequencing (and so can be used for differential expression analyses) and perhaps lowly expressed clock gene transcripts are swamped by such genes from elsewhere in the head and eye. Dissection of the brain specifically would perhaps have produced better evidence of clock gene expression. That this was not done was due to two considerations – firstly that the transcriptome produced for *Euphausia superba* used the entire head tissue and did not suffer from such issues, and secondly due to the shape of *Parhyale*'s brain and the nature of its surrounding tissue. As opposed to the relatively simple and rounded structure seen in *Drosophila*, the brain of *Parhyale* shows a number of protruding structures that are easily destroyed or lost during dissection, and it is furthermore surrounded by muscle anchored to the cuticle, making successful dissection of the brain alone and intact very difficult (Ramos, personal communication), particularly for mRNA extraction where time is of the essence. Perhaps on reflection it would have

been better to perform quick, imperfect dissections rather than use the entire head, but then if a putative central oscillator resided in a portion of the brain that was easy to lose by adopting such a technique it would have been self-defeating.

5.4.4.2 Cell specific expression

As well as the possibility that the genes are simply expressed at very low levels, it may also be that the clock genes are expressed only in very specific cells. In this scenario the majority of brain tissue does not produce clock gene transcripts either, enhancing the 'swamping' effect described above. There is in fact a small piece of evidence from *Eurydice pulchra* to support this idea, in that anti-*Eurydice* PERIOD immunoreactivity assays reveal expression in only 10 cells in the isopod brain (Zhang *et al.*, 2013). The truth of this in *Parhyale* awaits further research, but considering the small fragment of *Phperiod* found in both transcriptomes and its minimal expression data, it would not be surprising.

5.4.5 Future research

As revolutionary as RNA-seq is, should the next stage of this research focus on the discovery of the full coding sequences of the core circadian genes it might prove more productive to return to the approach that discovered a number of *Euphausia*'s orthologs. Degenerate PCR should not be necessary (unless one is unwilling to accept *timeless* and *cry1* are absent in *Parhyale*), as primers can be designed for *Phclock* and *Phperiod* using the fragments identified from the genome and transcriptome, followed by RACE extension. Alignment with orthologs suggests that a PCR based on genomic *Phclock* should produce a fragment approximately 1000 bp in length, and simply successfully doing so will answer one of the big questions posited by the results of this chapter. This was attempted without success prior to the submission of this thesis, but that should not be considered the final word on the matter. Should the final word deem that *Phclock* is in fact a genetic remnant that is not expressed, a yeast two-hybrid approach could be employed to pull down potential partners for PhBMAL1, and S2 cell work to assess their functionality.

The low level of clock gene expression seen here also suggests that studying gene expression across time-points might be better achieved using RT-PCR rather than

analysis of RNA-seq data, even with replicate data, as the probability of detecting differential expression decreases with gene expression level (Conesa *et al.*, 2016), while RT-PCR allows the researcher to focus on optimising an assay specific to that particular gene.

The other major issue outlined here is the location of clock gene expression. Antisera against PhBMAL1, PhCRY2 and PhPER could be raised and used in immunoreactivity assays, and for genes so far unidentified antibodies from other species can be attempted – for example using *Drosophila* anti-PDF to try and settle the issue of the presence or absence of PDH in *Parhyale*. *In situ* hybridisation (Rehm *et al.*, 2009b) and the ability to fuse GFP reporters to particular genes through the use of the CRISPR/Cas9 system (Serano *et al.*, 2016) would also be well employed in locating the sites of clock gene expression.

5.4.6   Conclusions

Two core circadian genes have been identified and verified for *Parhyale hawaiensis* through the assembly and mining of a *de novo* transcriptome and comparison with independent transcriptomic and genomic resources, along with a suite of regulatory and output genes. A further gene has been identified in fragmentary form in transcriptomic data and more extensively in the genome. Most intriguingly, the core gene *Phclock* cannot be detected in transcriptomic data and only in small fragmentary form in the genome and may not be expressed at all, while orthologs of *timeless* and *cryptochrome 1* have no supporting evidence in any resource. It is suggested that the detection of very low expression levels of particular circadian genes in the head transcriptome may be related to cell-specific expression patterns, and further research is required not only to sequence the remaining genes, if they exist, but to characterise their temporal and spatial expression patterns.

The very reason this species was investigated as a potential model organism for circadian research was its extensive molecular toolkit. In line with that other model organism *Drosophila melanogaster,* with its truncated CYC protein and absent CRY2, on first viewing the molecular clock of *Parhyale* is something of an oddity. Fortunately, it is an excellent organism in which to investigate oddities.

Chapter 6 General Discussion

6.1   Two circadian systems and their output

The preceding results chapters have documented investigations into the circadian systems of two crustaceans at both molecular and behavioural levels. In this chapter these findings will be considered in a broader context of ecological meaning and evolution and suggestions made for future research.

6.2   *Euphausia superba*

6.2.1   The molecular basis of *Euphausia's* clock

Chapter Two describes the cloning and characterisation of the core circadian genes of *Euphausia superba*, including the first work in this species describing how the associated proteins may interact to generate rhythmic output at the biochemical, physiological and behavioural level. To briefly recap: EsCLK and EsBMAL1 are required to drive E-box mediated gene expression, and EsPER, EsTIM and EsCRY2 show varying degrees of potency in repressing this activity individually. EsPER and EsTIM show moderate repressive capacity apart but act much more potently together, on a par with the strong repressor EsCRY2, and these results suggest *Euphausia* potentially has two negative limbs to its central feedback loop. Further to this, in Chapter Three a comprehensive suite of regulatory and putative clock controlled genes were identified and in many cases confirmed by PCR and Sanger sequencing.

The genes behind the core peptides have been shown to be expressed across the entire body from antennae to abdomen, and while further work is necessary to generate robust support there is some evidence of transcript cycling in *Escry1* (Chapter Two) and *Escry2* (Teschke *et al.*, 2011). Regarding the latter study, the authors identify and discuss a notably short endogenous period of ~18 hours. They note that Daan and Pittendrigh (1976) found that across various rodent species, short (<24 hour) or long (>24 hour) endogenous periods showed high amplitude phase response curves. If *Euphausia*'s apparently short endogenous period showed a similar high amplitude PRC, this might be informative about the nature of the underlying oscillator as those with large phase responses tend to be weak (slower to return to equilibrium after

perturbation), resulting in a capacity for a wide range of entrainment (Abraham *et al.*, 2010). Considering the extreme variation in photoperiod that *Euphausia* is subject to annually, this would not be surprising. The generation of a PRC for *Euphausia* to test the first stage of this logic, however, may prove difficult given the current lack of robust circadian phenomena that can be measured with relative ease.

6.2.2 Resulting rhythms

6.2.2.1 Chromatophores

Euphausia shows concentration and dispersal of erythrophore pigment on exposure to photosynthetically active radiation (PAR), a response thought to be employed to minimise UV damage, to which krill are particularly susceptible (Newman *et al.*, 1999). A circadian rhythm of chromatophore dispersal has been identified in *Eurydice pulchra* (Wilcockson *et al.*, 2011) and one might guess that the phenomenon is under similarly endogenous control in *Euphausia*, but on the current evidence this does not appear to be the case. Auerswald *et al.* (2008) found no circadian pattern, with low levels of dispersal at all timepoints (Figure 6.1).
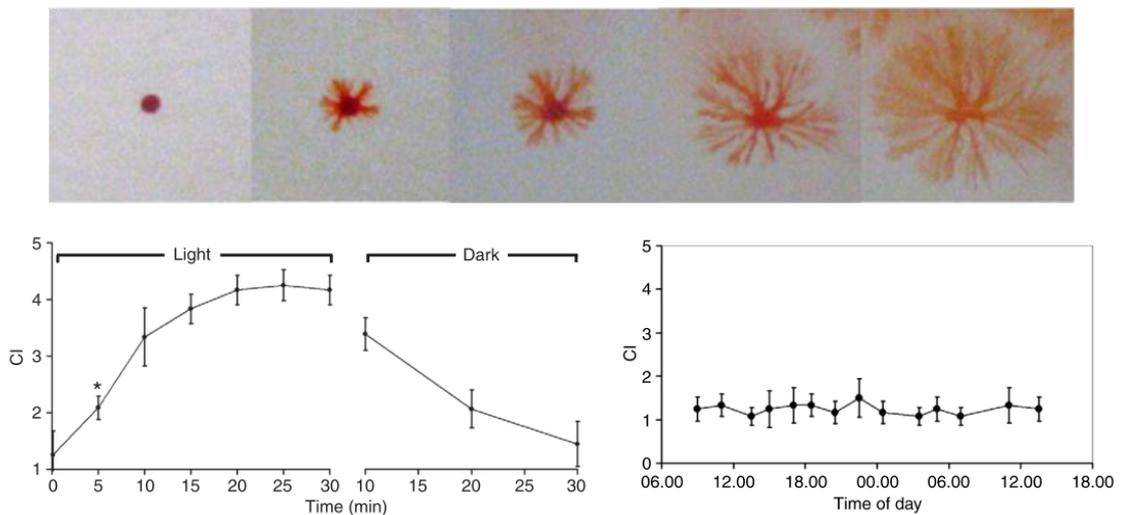
Figure 6.1: Top – representative of chromatophore index (CI 1 – 5) used to measure concentration and dispersion of *Euphausia superba* erythrophores. Bottom left – response to PAR exposure and return to darkness. Bottom right – dispersion over time in constant darkness. From Auerswald *et al* (2008).

Perhaps there has been no particularly strong selection pressure to link chromatophore patterns to the clock. It may be that an intertidal animal such as

Eurydice is under strong pressure to anticipate exposure to UV while, as a pelagic animal, *Euphausia* has a greater opportunity to avoid light through the simple expedient of moving down the water column – experimental evidence suggests that krill are willing to take such an opportunity when it is offered (Newman *et al.*, 2003). Further to this, schools of *Euphausia* are still sometimes seen near the surface during daylight hours, particularly when phytoplankton is abundant (Taki *et al.*, 2005), and five minutes of exposure to very low levels of PAR can induce erythrophore dispersal. Therefore *Euphausia* may employ only a reactive chromatophore system that is triggered as and when the animal chooses exposure. Two points regarding the study that failed to find a circadian connection should be noted, however. Firstly, it was conducted on isolated animals that had been kept in constant darkness for at least 24 hours prior to measurement, and it may be that a longer period of entrainment may produce a different result. Secondly, it was performed on winter-caught krill. Auerswald *et al.* (2008) did find a seasonal difference in the speed of chromatophore changes, with winter krill being slower to respond, and other polar animals have been found to be rhythmic only during the summer months (van Oort *et al.*, 2005). The possibility therefore remains that chromatophore concentration and dispersal in *Euphausia* may be seasonally circadian, as it were.

If this proves to be the case the identification and characterisation of *Euphausia*'s circadian genes will provide a solid basis for further investigation, particularly if RNAi can be established for this species and employed to knock down circadian gene expression and assess its impact on chromatophore rhythm.

6.2.2.2 Diel vertical migration

The DVM of *Euphausia* is not easy to characterise. As mentioned above it undergoes seasonal variation in strength and it is also impacted by hunger and satiety (Tarling and Johnson, 2006) to generate an overall complex pattern. But there is evidence that it may be under the control of the biological clock (Gaten *et al.*, 2008), with significant circadian, ultradian and 12-hour periodicities (considered part of a bimodal circadian component; Figure 6.2) identified using CLEAN spectral analysis. One of the challenges in assessing DVM – indeed, assessing behaviour in general – lies in allowing the animal to show its natural response, which in *Euphausia* represents vertical migration

measured in at least dozens of metres. For such an animal probably the only barrier regularly encountered in nature is the surface of the ocean and one wonders about its contentedness in a container 48 cm high and 15 cm in diameter. It is impressive to have gained evidence of rhythmicity in such an arena, but the difficulty in gathering and interpreting such data does not bode especially well for future experimental interventions employing the genetic knowledge described here. To quote from Gaten *et al.* (2010) "[from observing] pelagic crustaceans in aquaria, one could come to the naive conclusion that they lie at the bottom of the tank on their sides swimming in circles". Were RNAi successfully implemented in *Euphausia*, the issue arises as to whether or not we would even be able to discern its behavioural effects.
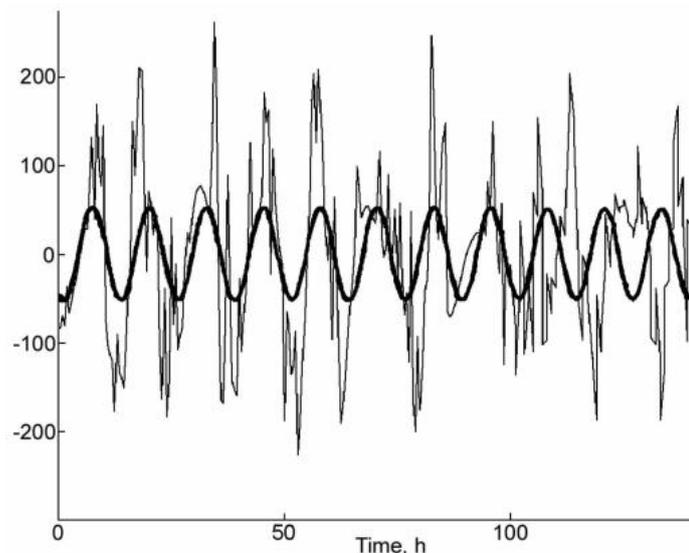
Figure 6.2: Bimodal (12.6 hour) circadian pattern of *Euphausia superba*. Chart shows relative activity over time for an individual in constant darkness, fitted to sinusoidal wave derived from the observed period. From Gaten *et al.* (2008).

One of the questions that arises regarding a species that experiences winters of almost constant darkness is how rhythms are maintained across the season – endogenous clocks can only do so much and it is common to see rhythmicity drift and dampen the longer they go without reentrainment from a *zeitgeber*. The answer may be moonlight. Last *et al.* (2016) found that as winter arrives and the days dramatically shorten the vertical migration patterns of Arctic zooplankton shift from a 24 hour DVM to a 24.8

hour lunar-day vertical migration (LDVM) driven by lunar altitude and thus irradiance, and furthermore observed a periodic suppression of these patterns tied precisely to the timing of the full moon. If and how such a phenomenon plays out in the Antarctic, and regarding *Euphausia* in particular, remains to be seen, as its relationship to light is still unclear. The diel vertical migration pattern is more pronounced during summer (Taki *et al.*, 2005) but when studying individuals they were not found to entrain to light at an intensity like that experienced at normal daylight depths (Gaten *et al.*, 2008) and it may be that light avoidance contributes to the extent of the summer DVM rather than robust entrainment. If they possess visual acuity in line with Arctic zooplankton it is possible that they might instead entrain to moonlight, and it would be interesting to repeat the experiment at much lower irradiance. It is also worth noting that a study into the vertical migration of the marine copepod *Acartia clausi* found that migration was induced by *changes* in light intensity, rather than its presence or absence (Johnson, 1938), with no movement seen if the light was held static at a particular intensity. The effect of a simulated natural light regime on *Euphausia*'s migration behaviour could also be illuminating.

6.2.2.3 Other assays

Recently a system was reported allowing the study of internal physiology of live krill through the development of a 'krill trap', a flow through system in which krill are held still for up to 7 hours without ill effect (Cox *et al.*, 2015), allowing the non-intrusive estimation of heart rate by video imaging. Should it prove possible to extend its use beyond 7 hours then there is a possibility it could be used in investigating circadian heart rate patterns as have been found in the lobster *Homarus americanus* (Chabot and Webb, 2008), or identifying other rhythmic physiological processes.

6.2.2.4 Seasonal variation

In section 2.4.3.2 various ideas were posited to explain the apparent ability of *Euphausia*'s clock to repress transcription via two separate pathways, one of which was the suggestion of distinct roles for each in maintaining circadian and seasonal rhythmicity. Capturing *Euphausia* in the Antarctic winter is a formidable undertaking and so our krill were taken only during the summer, and thus no comparative studies such as differential expression of the genes can be performed with those current

samples. Indeed a comprehensive approach would require some form of mutagenesis or knock out in live animals, such as to see the effects of clock ablation on the photoperiodic regulation of gene expression (Seear *et al.*, 2009). To get truly ambitious for a moment, *Euphausia* have been successfully maintained in aquaria for up to a year (Kawaguchi *et al.*, 2010) over which time sexual regression and re-maturation can be observed (Kawaguchi *et al.*, 2007), a phenomenon suggested to be controlled by an endogenous rhythm (Thomas and Ikeda, 1987). Furthermore reproduction appears to be induced by photoperiod cues (Hirano *et al.*, 2003). The technical challenge is daunting, but a long-term study of the responses of *Euphausia* clock mutants to these cues would go a long way towards deciding the importance of particular components to *Euphausia*'s annual rhythms.

6.3 *Parhyale hawaiensis*

6.3.1 A curious clock

The molecular clock of *Parhyale* warrants much further investigation. Gene expression appears to be low across all identified components to the extent that *Phperiod* is represented only by a small fragment in transcriptomic data and *Phclock* is not present at all, and may not even be expressed. In the current genome assembly, minimal evidence of that gene is seen, in highly fragmented form.

One idea, perhaps rather fanciful but worthy of consideration, is that laboratory *Parhyale* may have been subject to selection away from a typical circadian clock. A prior example, tellingly related to rhythmicity, is seen in certain strains of lab mice, which breed year round in comparison to seasonally reproductive wild types due to the loss of pineal-melatonin synthesis via mutations in two genes (Ebihara *et al.*, 1986). *Parhyale* (the model organism rather than the wild animal) has not only been reproducing for hundreds of generations in 12:12 luxury with predictably maintained conditions and food provision, but prior to this had spent an unknown amount of time subsisting on the waste filter of an aquarium, presumably a rather atypical environment in which robust circadian behaviour may have meant little. The answer to this theory lies in the collection and sequencing of genes from wild caught samples. As

a circumtropical species (Myers, 1985) this would also give the opportunity to investigate potential clines.

6.3.2 Going forward

The collection of wild samples might also enable productive testing of the hypothesis that *Parhyale* exhibits circa-tidal rhythmicity. While it seems unlikely that such rhythms have been maintained in the lab there is the possibility that they can be re-entrained with the right *zeitgeber*, be it agitation, immersion, hydrostatic pressure or changes in salinity. If nothing is seen in the lab animals, perhaps it will be more evident in wild strains, or in populations found in locations particularly conducive to circa-tidal behaviour.

Locomotor activity appears to be under the control of a circadian clock, despite the current molecular mystery, and further work is required to develop this assay. Akin to the section above regarding *Euphausia*, the key is in providing conditions in which the animal can exhibit natural behaviour. One possibly productive approach, considering its apparent photophobia, is the monitoring of emergence from, and return to, a covered 'den' with the use of an infra-red beam (Figure 6.3) or video monitoring. If food provision can be achieved without water spoilage this could also be used in conjunction with such a system to assess if feeding patterns are under endogenous control.
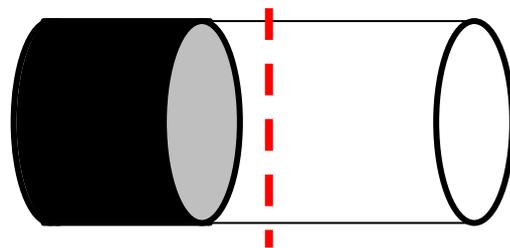
Figure 6.3: Emergence monitor. A proportion of a transparent plastic tube is covered in opaque black material to provide a region away from direct downwelling irradiance. An infra-red barrier passes across the tube near the mouth of the 'den', registering whenever the animal leaves or returns. Food provisioned at the far end could create a tension between perceived safety and food and encourage feeding rhythms.

Regarding the perceived photophobia/nocturnal behaviour observed in the lab, this too warrants further investigation. While it was in line with the maximum feeding rate data collected for wild-caught Australian specimens (Poovachiranon *et al.*, 1986), it does stand in contrast – if only slightly – to a study assessing the photopositive/negative responses of a small number of individuals from two populations in Lanzarote (Rodriguez, 1996). This found a photonegative response at 620 lux but a mild photopositivity at 3 lux and in one population a very mild photopositivity at 50 lux. As the locomotor activity experiments were conducted at ~300 lux, it may be worth testing varying degrees of irradiance to optimise the assay, particularly given the light-mediated suppression of a second peak apparent in some actograms.

It will also be worthwhile investigating the possibility of other assays. No obvious patterns of pigment migration in the eye or elsewhere on the body have been observed (data not shown) but further study, particular in the eye at the ultrastructural level, may reveal more subtle measures of rhythmicity. Finally, developmental studies have identified putative extra-retinal sensory/photoreceptor cells projecting to the brain (Extavour, 2005; Pavlopoulos and Averof, 2005), and these should be given particular attention in future; genetic or cell-ablation approaches may reveal such structures to be of importance to the *Parhyale* clock. Overall, there is compelling evidence that the further establishment of *Parhyale hawaiensis* as a model organism for chronobiology will be most productive.

6.3   Clock evolution

In *Daphnia pulex* (Tilden *et al.*, 2011), two species of copepod (Christie *et al.*, 2013a; Nesbit and Christie, 2014) and now in *Euphausia*, all the canonical components of the molecular clock – CLOCK, BMAL1, PERIOD, TIMELESS, a *Drosophila*-like CRY1 and a mammalian CRY2 – are present, while *Eurydice pulchra* lacks only a CRY1 (Zhang *et al.*, 2013). On the assumption that evidence of *Phclock* expression will be found in due course given its presence in the genome, *Parhyale hawaiensis* offers a third type of crustacean clock. This system, lacking TIM and CRY1, is seen elsewhere in vertebrates and hymenopterans.

The repeated emergence of this particular system is interesting. TIM and CRY1 interact directly in a light-dependent manner in *Drosophila* to release the repressive effect of the PER:TIM dimer (Rosato *et al.*, 2001) and have also been linked with seasonal diapause (Sandrelli *et al.*, 2007), and this close association may explain the repeated instances in which the loss of one is coincident with the loss of the other. It may even be that this is an example of convergent molecular evolution, and a particular selection pressure encourages the generation of this molecular system in much the same way that the fusiform shapes of various swimming species arise independently. What might cause the evolution of particular type of clock is a big question far beyond the scope of this thesis, though that did not prevent a brief flirtation with a flawed hypothesis, outlined as follows.

Hymenoptera contains numerous examples of eusocial species, and while there is no evidence of eusociality in *Parhyale* it is found in dense populations of up to 7,000 animals per m^2 (Poovachiranon *et al.*, 1986). On the hypothesis that population density or some level of sociality may have some influence on clock evolution, the genomes of two species of the order Blattodea were searched for clock genes. In the eusocial termite *Zootermopsis nevadensis* (Terrapon *et al.*, 2014) once again no evidence was found of *timeless* or *cry1* and it appears to be the same system as is found in hymenopterans. The cockroach *Blattella germanica*, on the other hand, shows evidence of all core components (Murali *et al*, unpublished; see genome GCA_000762945.1 at NCBI). This clock type is clearly not a direct cause or inevitable consequence of eusociality, as there are plenty of flaws that undermine this idea in such a naive form – *Euphausia* is also a highly social animal living in enormously dense swarms yet has all clock components; there are many hymenopteran species that are solitary; and no explanation is offered regarding the evolution of the same system in vertebrates, to name just three. But with further research perhaps a common thread can be found, or the suggestion that it means anything can be definitively put to bed as noise from a limited dataset.

Such a diverse array of molecular clocks have now been identified that it is perhaps time to stop speaking of the mammalian clock, the insect clock, or the crustacean clock – the former is clearly not unique to mammals, and the diversity of insect systems is

now almost matched by Crustacea. It may instead be better to speak of clock types that are found repeatedly across the clades (Table 6.1). While there is little diversity to be seen in the chordates, it should not also be automatically assumed that diversity is a feature unique to Pancrustacea (though this may be true, and is itself an intriguing thought). Further study of other arthropods, annelids, molluscs and other phyla is required to resolve this aspect of the evolution of the clock.

The many genome projects reaching completion over recent years are beginning to offer a broad view into the beginnings of the clock. Genes have been identified in the sponge *Amphimedon queenslandica* encoding bHLH proteins thought to be orthologous to CLOCK and BMAL1, though the case is not clear-cut as the putative CLOCK shows fairly weak sequence similarity and was assigned to this family mostly on the basis of a conserved PAS domain, while the putative BMAL1 acts as the outgroup to both ARNT and BMAL families (Simionato *et al.*, 2007). Two cryptochromes have been found in that species and one in another poriferan, *Suberites domuncula* (Müller *et al.*, 2010; Rivera *et al.*, 2012), but neither TIMELESS nor PERIOD are found in sponges. Neither are they seen in cnidarians, although here both *Drosophila* and mammalian-like CRYs have been identified (Reitzel *et al.*, 2010; Brady *et al.*, 2011; Hoadley *et al.*, 2011; Shoguchi *et al.*, 2013). Looking at the protostomes, the full set of core circadian peptides is found in the polychaete worm *Platynereis dumerilii* (Zantke *et al.*, 2013) as well as various arthropods, and while TIMELESS is lacking in the chordates and PERIOD is lacking in the sea urchin *Strongylocentrotus purpuratus*, the deuterostomes as a whole show evidence of all components. This suggests that PERIOD and TIMELESS arose early in the Bilateria and were subsequently lost in various lineages, and the most parsimonious hypothesis is that the ancestral Bilaterian clock possessed all the core components (Figure 6.4).

A

Phylum	Subphylum	Class	Order	Species	CLOCK	BMAL/ CYCLE	TIMELESS	PERIOD	CRY-d	CRY-m	Bilateria Type
Arthropoda	Hexapoda	Insecta	Diptera	*Aedes aegypti*	Y	Y	Y	Y	Y	Y	I
				Anopheles gambiae	Y	Y	Y	Y	Y	Y	I
				Drosophila melanogaster	Y	Y	Y	Y	Y	N	II
				Musca domestica	Y	Y	Y	Y	Y	N	II
			Lepidoptera	*Antheraea pernyi*	Y	Y	Y	Y	Y	Y	I
				Bombyx mori	Y	Y	Y	Y	Y	Y	I
				Danaus plexippus	Y	Y	Y	Y	Y	Y	I
			Hymenoptera	*Apis mellifera*	Y	Y	N	Y	N	Y	III
				Nasonia vitripennis	Y	Y	N	Y	N	Y	III
				Solenopsis invicta	Y	Y	N	Y	N	Y	III
			Coleoptera	*Tribolium castaneum*	Y	Y	Y	Y	N	Y	IV
			Hemiptera	*Acyrthosiphon pisum*	Y	Y	Y	Y	Y	Y	I
			Orthoptera	*Gryllus bimaculatus*	Y	Y	Y	Y	Y	Y	I
			Blattodea	*Zootermopsis nevadensis*	Y	Y	N	Y	N	Y	III
				Blattella germanica	Y	Y	Y	Y	Y	Y	I
	Crustacea	Malacostraca	Euphausiacea	*Euphausia superba*	Y	Y	Y	Y	Y	Y	I
			Isopoda	*Eurydice pulchra*	Y	Y	Y	Y	N	Y	IV
			Amiphipoda	*Parhyale hawaiensis*	Y	Y	N	Y	N	Y	III
		Maxillopoda	Calanoida	*Calanus finmarchicus*	Y	Y	Y	Y	Y	Y	I
			Harpacticoida	*Tigriopus californicus*	Y	Y	Y	Y	Y	Y	I
		Branchiopoda	Cladocera	*Daphnia pulex*	Y	Y	Y	Y	Y	Y	I
	Chelicerata	Arachnida	Araneae	*Stegodyphus mimosarum*	Y	Y	Y	Y	Y	Y	I

Table continues overleaf

194

B

Phylum	Subphylum	Class	Order	Species	CLOCK	BMAL/ CYCLE	TIMELESS	PERIOD	CRY-d	CRY-m	Bilateria Type
Annelida		Polychaeta	Phyllodocida	*Platynereis dumerilii*	Y	Y	Y	Y	Y	Y	I
Nematoda		Chromadorea	Rhabditida	*Caenorhabditis elegans*	Y	Y	Y	Y	N	N	V
Chordata	Vertebrata	Mammalia	Rodentia	*Mus musculus*	Y	Y	N	Y	N	Y	III
		Aves	Galliformes	*Gallus gallus*	Y	Y	N	Y	N	Y	III
		Amphibia	Anura	*Xenopus laevis*	Y	Y	N	Y	N	Y	III
	Cephalochordata	Leptocardii	Amphioxiformes	*Branchiostoma floridae*	Y	Y	N	Y	N	Y	III
Echinodermata		Echinoidea	Echinoida	*Strongylocentrotus purpuratus*	Y	Y	Y	N	Y	Y	VI
Cnidaria		Anthozoa	Scleractinia	*Acropora millepora*	Y	Y	N	N	Y	Y	-
				Acropora digitifera	Y	Y	N	N	Y	Y	-
				Favia fragum	Y	Y	N	N	Y	Y	-
			Actiniaria	*Nematostella vectensis*	Y	Y	N	N	Y	Y	-
Porifera		Demospongiae	Haplosclerida	*Amphimedon queenslandica*	?	?	N	N	Y	N	-
			Hadromerida	*Suberites domuncula*	?	?	N	N	Y	N	-

Table 6.1: Diversity of clock types in A) Arthropoda (see previous page) and B) other animal phyla, based on published literature and genome searches. It is suggested here that the last common ancestor of all bilaterians possessed a clock with all the core molecular components identified across all modern descendents, some of which are subsequently lost in certain taxa to generate bilaterian clock types (no 'type' therefore shown for Cnidaria and Porifera). Vertebrate and echinoderm data from Rubin *et al.* (2006). Insect data from Tomioka and Matsumoto (2015) with the exception of Blattodea. Crustacean data from this thesis, Zhang *et al*, (2013); Tilden *et al*, (2011); Nesbit and Christie, (2014); Christie *et al.* (2013a). *Stegodyphus mimosarum* and *Branchiostoma floridae* data mined from genomes (Sanggaard *et al.*, 2014; Putnam *et al.*, 2008). *Caenorhabditis elegans* data from Romanowski *et al.* (2014). All other sources see text. Note that CLOCK and CYCLE/BMAL1 in the Porifera are putative (marked ?) due to relatively weak sequence similarity or clustering with more than one bHLH family (Simionato *et al.*, 2007). Putative CLOCK and CYCLE/BMAL1 are assumed for *Suberites domuncula* based on the evidence from *Amphimedon queenslandica*. CRY-d – *Drosophila*-like. CRY-m – mammalian-like.

Clade	No. of representatives	Bilaterian clock types	Components not found
Porifera	2	-	PER, TIM, CRY2
Cnidaria	4	-	PER, TIM
Vertebrata	3	III	TIM, CRY1
Branchiostomiformes	1	III	TIM, CRY1
Echinodermata	1	VI	PER
Insecta	14	I, II, III, IV	-
Crustacea	6	I, III, IV	-
Arachnida	1	I	-
Nematoda	1	V	CRY1, CRY2
Annelida	1	I	-
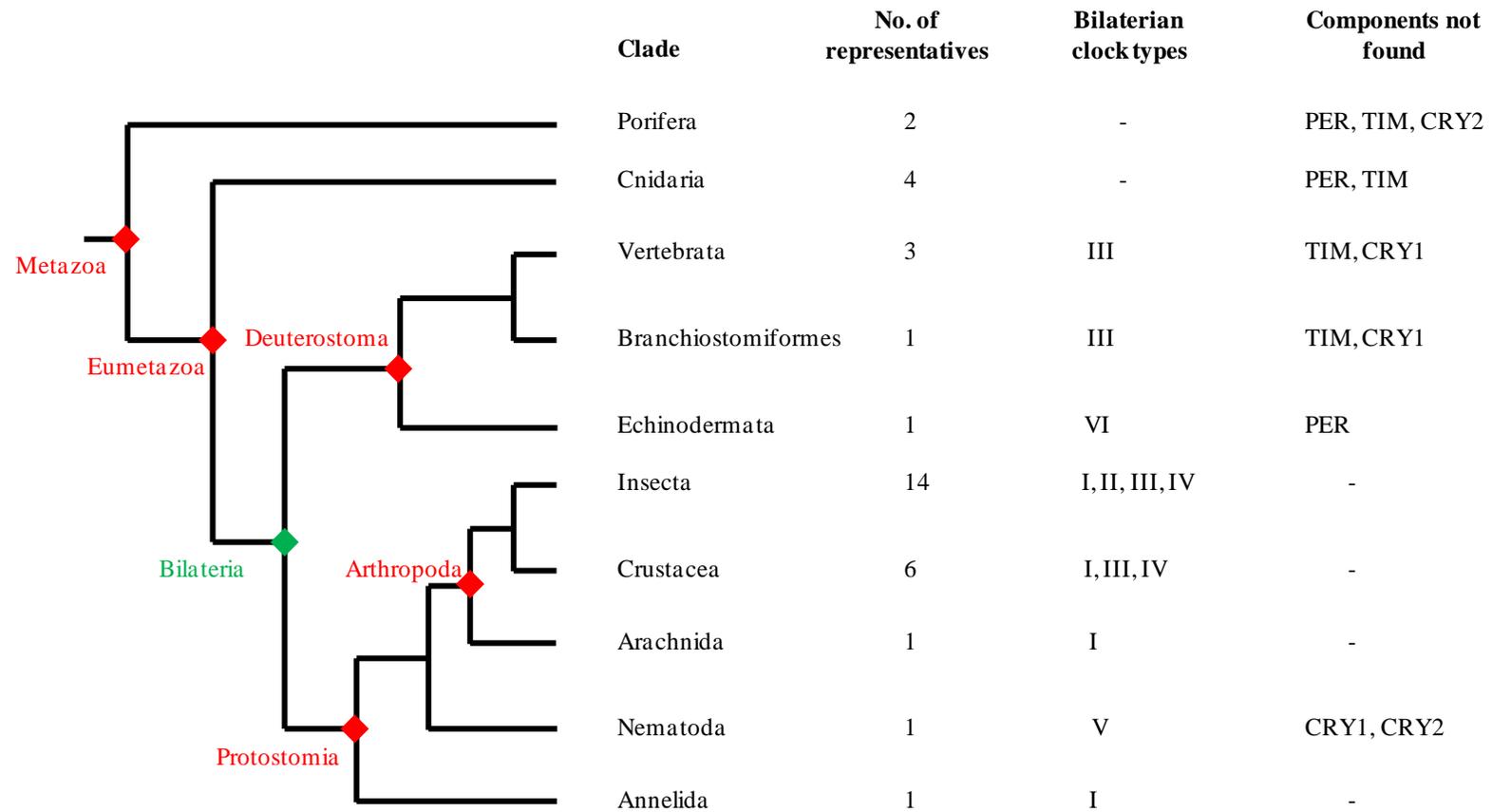
Figure 6.4: Phylogeny of clades shown in Table 6.1. PER and TIM are not seen until after the divergence of Bilateria while other components are found in Porifera and Cnidaria. All components are found in Deuterostomes as a group, and in select insects, arachnids, crustaceans and annelids, implicating Bilateria as the point at which the 'complete' modern clock evolved (marked in green). Phylogenetic tree redrawn and adapted from Simionato *et al.* (2007).

6.4   Assembly of *de novo* transcriptomes for non-model organisms

Finally, in Chapter Three a pipeline for *de novo* assembly of non-model organism transcriptomes was described. The development of this pipeline was conducted with the aim of maximising opportunities for gene identification by employing the unique strengths of multiple assemblers and varying *k*-mer parameters. This was shown to produce more complete and accurate assembly of transcripts in the *Euphausia superba* transcriptome database when compared to a single-run transcriptome, and while there where initial concerns about the transcriptome produced using the same method for *Parhyale hawaiensis* those results seem more likely to stem from the nature of that animal's circadian gene expression patterns – comparison to an independently assembled transcriptome and the newly sequenced genome confirmed the accuracy of the contigs that were found and that the lack of certain genes was not likely to be due to poor assembly. One of the strongest performers was a recent tool named Bridger (Chang *et al.*, 2015) and as new assemblers – and new methods of assembly – arise it will be good practice to assess and include the best performers in the pipeline.

An unavoidable consequence of such an approach is a high level of duplicate contigs even after processing the combined transcriptome to reduce redundancy, and this is certain to have an effect on downstream processes. A solution was offered in the form of the use of objective read mapping metrics from the recently developed software Transrate (Smith-Unna *et al.*, 2015) to select a single best assembly from those generated as a representative transcriptome for further procedures such as differential expression analysis. In this way, then, the recommended approach is to generate two transcriptomes – one a comprehensive resource for identification of genes of interest, and the other a 'working' transcriptome upon which downstream analyses can be conducted.

6.5   Concluding remarks

The circadian systems of two crustacean species have been identified and characterised beyond the level of sequence data, both in the process generating some surprising results. The clock of *Euphausia superba* appears to have the unique ability to

strongly repress the transcriptional activation activity of its central heterodimer via two separate pathways that reflect the workings of both the *Drosophila* and *Mus* orthologous systems. *Parhyale hawaiensis* meanwhile lacks the core components TIMELESS and CRY1, and in this it resembles a type of clock that is also found in vertebrates, hymenopterans and – to our knowledge explicitly identified here for the first time – the eusocial termite *Zootermopsis nevadensis*. In the expression data of regulatory genes derived from the *Parhyale* transcriptome, a bimodal pattern of expression over 24 hours is seen that could be linked to the endogenous locomotor activity pattern detailed herein, with a significant 12 hour periodicity regularly identified in CLEAN spectral analysis. Given its basic requirements for lab culture, year-round reproduction, short generation time and extensive molecular toolkit, *Parhyale* is an exciting opportunity for crustacean circadian research.

Through analysis of the clock genes identified in various species it has further been determined that features such as the poly-Q of many CLOCK proteins is likely an ancestral feature, and further support generated for the hypothesis that the BCTR of BMAL1/CYC proteins is also ancestral. It is also noted that the circadian systems of Crustacea are almost as diverse as those of Insecta; future research will determine if this is a feature of Pancrustacea, or if clock diversity is the rule rather than the exception.

Appendix I – accessions used for generation of phylogenetic trees

Protein	Accession	Species
CYCLE/BMAL1	NP_524168	*Drosophila melanogaster*
	EFX90169	*Daphnia pulex*
	BAJ16354	*Thermobia domestica*
	NP_001107795	*Tribolium castaneum*
	NP_001164574	*Acyrthosiphon pisum*
	BAF35030	*Athalia rosae*
	AFV39705	*Pacifastacus leniusculus*
	AAF64394	*Danio rerio*
	AAL98706	*Gallus gallus*
	NP_031515	*Mus musculus*
	NP_001123206	*Ovis aries*
	AAW80970	*Xenopus laevis*
	AAR14937	*Antheraea pernyi*
	EHJ64590	*Danaus plexippus*
	ABI21880	*Lutzomyia longipalpis*
	AEX32872	*Aedes aegypti*
	XP_556301	*Anopheles gambiae*
	XP_001121441	*Apis mellifera*
	AGV28715	*Eurydice pulchra*
	XP_008215795	*Nasonia vitripennis*
	BAN28450	*Gryllus bimaculatus*
	JW523145	*Tigriopus californicus*
	GAXK01131751	*Calanus finmarchicus*
	AGX93014	*Platynereis dumerilii*
CLOCK	NP_523964	*Drosophila melanogaster*
	BAJ16353	*Thermobia domestica*
	AAX44045	*Macrobrachium rosenbergii*
	EFX79971	*Daphnia pulex*
	AAC53200	*Mus musculus*
	AAR14936	*Antheraea pernyi*
	NP_571032	*Danio rerio*
	XP_315720	*Anopheles gambiae*
	NP_001083854	*Xenopus laevis*
	NP_989505	*Gallus gallus*
	ABV71922	*Ovis aries*
	EHJ69324	*Danaus plexippus*
	EFA01240	*Tribolium castaneum*
	BAM76759	*Gryllus bimaculatus*
	NP_001164531	*Acyrthosiphon pisum*

	XP_394233	*Apis mellifera*
	XP_001599257	*Nasonia vitripennis*
	AFV39704	*Pacifastacus leniusculus*
	AGV28720	*Eurydice pulchra*
	AKN63486	*Lutzomyia longipalpis*
	GAXK01092177	*Calanus finmarchicus*
	AGX93013	*Platynereis dumerilii*
PERIOD	NP_525056	*Drosophila melanogaster*
	EFX76293	*Daphnia pulex*
	AAN02439	*Blattella germanica*
	BAG48878	*Gryllus bimaculatus*
	NP_001152839	*Mus musculus*
	NP_001011596	*Apis mellifera*
	EFA04566	*Tribolium castaneum*
	EHJ74075	*Danaus plexippus*
	NP_001025354	*Danio rerio*
	AAK97374	*Bulla gouldiana*
	Q17062	*Antheraea pernyi*
	NP_001079172	*Xenopus laevis*
	AGV28714	*Eurydice pulchra*
	ALC74274	*Nephrops norvegicus*
	XP_011977904	*Ovis aries*
	NP_989593	*Gallus gallus*
	XP_008209243	*Nasonia vitripennis*
	GAXK01127710	*Calanus finmarchicus*
	JW535312	*Tigriopus californicus*
TIMELESS/TIMEOUT	AAR15505	*Danaus plexippus*
	BAB85487	*Neobelleria bullata*
	AAF66996	*Antheraea pernyi*
	ABW71828	*Chymomyza costata*
	EFA04644	*Tribolium castaneum*
	AAY40757	*Aedes aegypti*
	BAJ16356	*Gryllus bimaculatus*
	NP_001157553	*Mus musculus*
	XP_002934575	*Xenopus laevis*
	XP_006565495	*Apis mellifera*
	NP_001265529	*Danio rerio*
	NP_722914	*Drosophila melanogaster*
	EFX87311	*Daphnia pulex*
	AGV28716	*Eurydice pulchra*
	EEZ99220	*Tribolium castaneum*
	EHJ76705	*Danaus plexippus*
	AAF54908	*Drosophila melanogaster*
	GAXK01195225	*Calanus finmarchicus*

Appendix II – sample commands for assembling a multi-*k*-mer, multi-assembler transcriptome.

Software links:

http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

http://www.usadellab.org/cms/?page=trimmomatic

http://cutadapt.readthedocs.io/en/stable/installation.html

https://github.com/trinityrnaseq/trinityrnaseq/releases

http://www.bcgsc.ca/platform/bioinfo/software/trans-abyss

https://sourceforge.net/projects/rnaseqassembly/files/?source=navbar

http://sourceforge.net/projects/soapdenovotrans/files/SOAPdenovo-Trans/

http://hibberdlab.com/transrate/installation.html

https://github.com/weizhongli/cdhit

https://sourceforge.net/projects/bbmap/

https://transdecoder.github.io/

Quality control:

Concatenate all reads in two paired read files. This command adds /1 or /2 to NextSeq reads, a required tag for some assemblers.

```
cat *_R1_*.fastq | sed '/^@/ s/$/\/1/' > all_1.fastq

cat *_R2_*.fastq | sed '/^@/ s/$/\/2/' > all_2.fastq
```

Trim NextSeq adapters and remove low quality reads

```
cutadapt    -a    GATCGGAAGAGC    -A    GATCGGAAGAGC    -o
cutadapt.1.fastq    -p    cutadapt.2.fastq    all_1.fastq
all_2.fastq
```

```
trimmomatic    PE    -threads    16    cutadapt.1.fastq
cutadapt.2.fastq    readfile1.fastq    readfile1_unpair.fastq
readfile2.fastq readfile2_unpair.fastq AVGQUAL:20 MINLEN:72
```

Assembly: De Brujin graph assemblers are computationally demanding. A typical resource request to ALICE to run one assembly would entail 500 Gb of RAM and 16 processors. Note that the Butterfly stage of Trinity requests a certain amount of RAM per processor and can crash the job if the total requested is higher than that provided as a whole. It is therefore good practice to limit the number of processors in this stage to ensure it remains below that limit (in the example below, Butterfly is limited to four processors). The following all generate 25 *k*-mer assemblies.

```
Bridger.pl    --seqType    fq    --left    readfile1.fq    --right
readfile2.fq -k 25 --CPU 16 --out /outputpath/outputfolder/
--clean
```

```
Trinity --seqType fq --JM 400G --left readfile1.fq --right
readfile2.fq --output /outputpath/outputfolder/ --CPU 16 --
min_contig_length 200 --bflyCPU 4
```

```
transabyss    --pe    readfile1.fq    readfile2.fq    --outdir
/outputpath/outputfolder/ --name transabyss21 --length 200
--threads 16 -k 25
```

```
SOAPdenovo-Trans-127mer    all    -s    euphausia.config    -o
/outputpath/outputfolder/ -R -K 25 -L 200 -p 16
```

Merge assembler-specific assemblies

```
transabyss-merge    assembly21.fasta    assembly51.fasta
assembly91.fasta --mink 21 --maxk 91 --prefixes k21. k51.
k91. --out transabyss.merge.fasta
```

Or, with all assemblies in one folder, first modify the headers to specify the k-mer of each assembly to avoid conflicting IDs after merging

```
sed -i 's/>/>k71./g' assembly71.fasta
```

Then combine the assemblies

```
cat *.fasta > merged.fasta
```

And finally remove duplicates at 100% with either bbmap or CD-HIT

```
dedupe.sh in=merged.fasta out=merged_dedupe.fasta
outd=duplicates.fasta overwrite=t absorbrc=t
```

```
cd-hit-est -i merged.fasta -o merged_cdhit.fasta -c 1 -n 10
-M 60000
```

Transrate

Run for each assembler-specific merged assembly. Transrate is typically as resource hungry as the assemblers themselves.

```
transrate --assembly merged.fasta --left readfile1.fq --
right readfile2.fq --threads 8 -o /outputpath/outputfolder/
```

Take the good contigs from each run, combine and deduplicate again as above. Run Transrate on this combined file and take the good contigs to create the **total assembly**.

Coding contigs

Create the peptide assembly as follows (requires Trinity).

```
perl TransDecoder.LongOrfs -t totalassembly.fasta
```

```
perl TransDecoder.Predict -t totalassembly.fasta
```

Finally, identify the **coding assembly** contigs from the **peptide assembly**. Make a list of the contigs appearing in the latter, and extract them from a BLAST database (see below) made from the total assembly using this command.

```
blastdbcmd -db totalassembly.fasta.db -dbtype nucl -
entry_batch codingcontigs.txt -outfmt %f -out
codingassembly.fasta
```

Building a BLAST database and querying it

The following commands make a nucleotide database and then query it with protein sequences. For nucleotide queries use blastn rather than tblastn.

```
makeblastdb -dbtype nucl -in totalassembly.fasta -out
totalassembly.fasta.db -parse_seqids
```

```
tblastn -query querycontigs.fasta -db
totalassembly.fasta.db -out querycontigs.output -html -
evalue 0.001
```

Annotation

Use the following commands to generate BLAST output for the coding and peptide assemblies in an XML format that can be used in BLAST2GO

```
blastx -db uniprotkb_arthropoda__mod_db -outfmt 5 -evalue
1e-6 -word_size 3 -show_gis -max_target_seqs 1 -num_threads
16 -out blast.xml -query codingassembly.fasta
```

```
blastp -db uniprotkb_arthropoda__mod_db -outfmt 5 -evalue
1e-6 -word_size 3 -show_gis -max_target_seqs 1 -num_threads
16 -out pep.blast.xml -query peptideassembly.pep
```

Appendix III – Thermal profiles

Profile	Temp (°C)	Duration (mm:ss)	
1	94	01:00	
	45	01:00	x5
	70	01:00	
	94	01:00	
	50	01:00	x35
	70	01:00	
	72	10:00	
2	94	01:00	
	94	01:00	
	54	01:00	x30
	72	01:00	
	72	01:00	
3	94	05:00	
	94	01:00	
	65	01:00	x30
	68	01:00	
	68	10:00	
4	94	05:00	
	94	01:00	x5
	72	01:00	
	94	01:00	x5
	70	01:00	
	94	01:00	
	65	01:00	x25
	68	01:00	
	68	10:00	
5	94	05:00	
	94	01:00	
	65	01:00	x25
	68	01:00	
	68	10:00	
6	94	05:00	
	94	01:00	
	65	01:00	x25
	68	01:30	
	68	10:00	
7	94	05:00	
	94	01:00	x5
	72	01:00	
	94	01:00	x5
	70	01:00	
	94	01:00	
	65	01:00	x25
	68	01:00	
	68	10:00	
8	94	05:00	
	94	01:00	x5
	72	01:30	
	94	01:00	x5
	70	01:30	
	94	01:00	
	65	01:00	x25
	68	01:30	
	68	10:00	

9	94	05:00	
	94	01:00	
	65	01:00	x25
	68	01:00	
	68	10:00	

10	94	05:00	
	94	01:00	
	65	01:00	x25
	68	01:30	
	68	10:00	

11	94	05:00	
	94	01:00	x5
	72	02:00	
	94	01:00	x5
	70	02:00	
	94	01:00	
	65	01:00	x25
	68	02:00	
	68	10:00	

12	94	05:00	
	94	01:00	
	65	01:00	x25
	68	02:00	
	68	10:00	

13	94	05:00	
	94	01:00	
	55	01:00	x35
	72	02:30	
	72	10:00	

14	94	05:00	
	94	01:00	
	60	01:00	x35
	72	02:30	
	72	10:00	

15	94	05:00	
	94	01:00	
	60	01:00	x30
	72	01:00	
	72	10:00	

17	94	05:00	
	94	01:00	
	55	01:00	x30
	72	01:00	
	72	10:00	

18	95	03:00	
	94	00:30	
	53	00:30	x35
	72	01:00	
	72	02:00	

19	95	03:00	
	94	00:30	
	55	00:30	x35
	72	01:00	
	72	02:00	

20	95	03:00	
	94	00:30	
	50	00:30	x35
	72	01:00	
	72	02:00	

21	94	05:00	
	94	01:00	} x5
	72	01:30	
	94	01:00	} x5
	70	01:30	
	94	01:00	
	65	01:00	} x25
	68	01:30	
	68	10:00	

22	94	05:00	
	94	01:00	} x5
	72	03:00	
	94	01:00	} x5
	70	03:00	
	94	01:00	
	65	01:00	} x25
	68	03:00	
	68	10:00	

23	94	05:00	
	94	01:00	
	65	01:00	} x25
	68	03:00	
	68	10:00	

24	98	00:30	
	98	00:10	
	varying	00:30	} x35
	68	02:00	
	68	10:00	

25	98	00:30	
	98	00:10	
	varying	00:30	} x35
	68	varying	
	68	10:00	

Appendix IV – *Estimeless* tissue expression

As noted in section 2.4.3.1, the expression of *Estimeless* appears to be very low in certain tissues, particularly the thorax but also the antennae and abdomen. Often the reactions would produce no clear band for one or more of these tissues and it required multiple efforts to confirm expression in all of them. The image used in Figure 2.7 was the best in terms of showing evidence of expression in all tissues, but is marred by unrelated genomic DNA contamination in the water control. This was considered acceptable as the *Estimeless* band is not present in this lane, indicating that the contamination was not of krill origin nor capable of producing a false positive if spread to other reactions. Nevertheless, further evidence is provided below of the expression of Estimeless in all tissues.
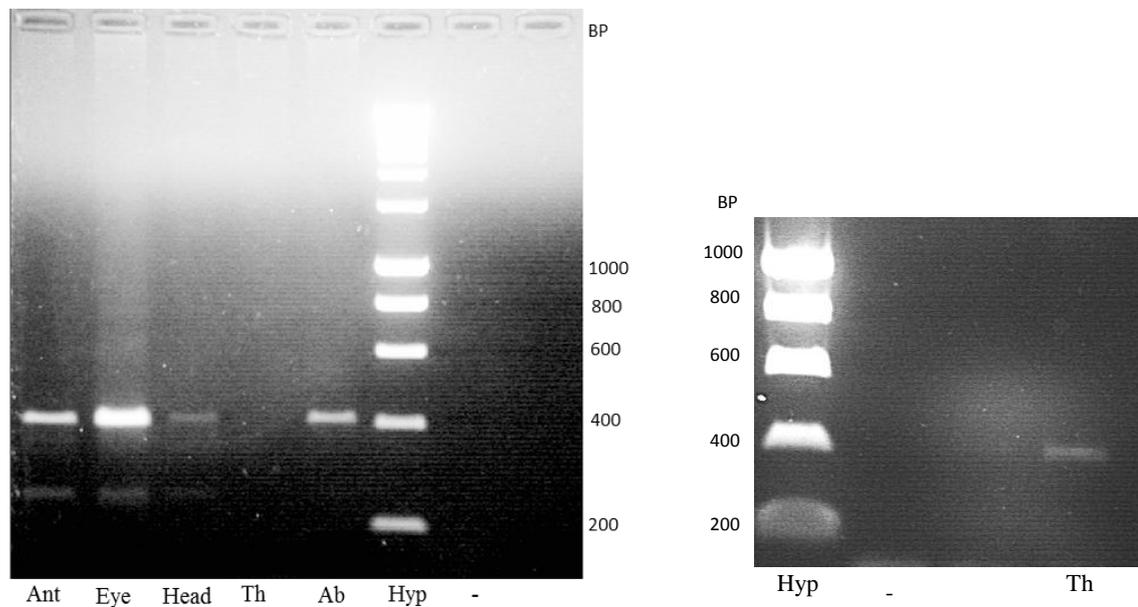
Above left shows expression in all tissues but the thorax, while above right shows expression in the thorax. Both images have an uncontaminated water control "-".

References

Abraham, U., Granada, A.E., Westermark, P.O., Heine, M., Kramer, A., Herzel, H. (2010) Coupling governs entrainment range of circadian clocks. *Molecular Systems Biology*. **6**(438), 438.

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

W., Zhou, X., Zhu, S., Zhu, X., Smith, H.O., Gibbs, R.A., Myers, E.W., Rubin, G.M., Venter, J.C. (2000) The genome sequence of *Drosophila melanogaster*. *Science*. **287**(5461), 2185–95.

Aguzzi, J., Chiesa, J.J. (2005) Cardiac activity of *Nephrops norvegicus* (Decapoda: Nephropidae): The relationship between circadian and ultradian rhythms. *Journal of Crustacean Biology*. **25**(4), 577–584.

Aguzzi, J., Costa, C., Menesatti, P., García, J.A., Chiesa, J.J., Sardà, F. (2009) Monochromatic blue light entrains diel activity cycles in the Norway lobster, *Nephrops norvegicus* (L.) as measured by automated video-image analysis. *Scientia Marina*. **73**(4), 773–783.

Aguzzi, J., Sardá, F. (2007) Biological rhythms in the marine environment: The Norway lobster as a case study. *Contributions to Science*. **3**(4), 493–500.

Akiyama, T. (1997) Tidal Adaptation of a Circadian Clock Controlling a Crustacean Swimming Behavior. *Zoological Science*. **14**(6), 901–906.

Akten, B., Jauch, E., Genova, G.K., Kim, E.Y., Edery, I., Raabe, T., Jackson, F.R. (2003) A role for CK2 in the *Drosophila* circadian oscillator. *Nature Neuroscience.* **6**(3), 251–257.

Alheit, J., Naylor, E. (1976) Behavioural basis of intertidal zonation in *Eurydice pulchra* Leach. *Journal of Experimental Marine Biology and Ecology*. **23**(2), 135–144.

Allada, R., White, N.E., So, W.V., Hall, J.C., Rosbash, M. (1998) A Mutant *Drosophila* Homolog of Mammalian *Clock* Disrupts Circadian Rhythms and Transcription of *period* and *timeless*. *Cell*. **93**(5), 791–804.

Alonzo, S.H., Mangel, M. (2001) Survival strategies and growth of krill: Avoiding predators in space and time. *Marine Ecology Progress Series*. **209**, 203–217.

Arechiga, H., Rodriguez-Sosa, L. (1998) Circadian clock function in isolated eyestalk tissue of crayfish. *Proceedings of the Royal Society B: Biological Sciences*. **265**(1408), 1819–1823.

Atkinson, A., Siegel, V., Pakhomov, E., Rothery, P. (2004) Long-term decline in krill

stock and increase in salps within the Southern Ocean. *Nature*. **432**(7013), 100–3.

Auerswald, L., Freier, U., Lopata, A., Meyer, B. (2008) Physiological and morphological colour change in Antarctic krill, Euphausia superba: a field study in the Lazarev Sea. *Journal of Experimental Biology*. **211**(Pt 24), 3850–3858.

Ayala, F.J., Rzhetsky, A., Ayala, F.J. (1998) Origin of the metazoan phyla: Molecular clocks confirm paleontological estimates. *Proceedings of the National Academy of Sciences*. **95**(2), 606–611.

Barnard, J.L. (1965) Marine Amphipoda of atolls in Micronesia. *Proceedings of the United States National Museum*. **117**(3516), 459–551.

Beckwith, E.J., Lelito, K.R., Hsu, Y.-W.A., Medina, B.M., Shafer, O., Ceriani, M.F., de la Iglesia, H.O. (2011) Functional Conservation of Clock Output Signaling between Flies and Intertidal Crabs. *Journal of Biological Rhythms*. **26**(6), 518–529.

Benito, J., Zheng, H., Hardin, P.E. (2007) PDP1 Functions Downstream of the Circadian Oscillator to Mediate Behavioral Rhythms. *Journal of Neuroscience*. **27**(10), 2539–2547.

Benna, C., Bonaccorsi, S., Wülbeck, C., Helfrich-Förster, C., Gatti, M., Kyriacou, C.P., Costa, R., Sandrelli, F. (2010) *Drosophila timeless2* is required for chromosome stability and circadian photoreception. *Current biology*. **20**(4), 346–52.

Benna, C., Scannapieco, P., Piccin, a., Sandrelli, F., Zordan, M., Rosato, E., Kyriacou, C.P., Valle, G., Costa, R. (2000) A second *timeless* gene in *Drosophila* shares greater sequence similarity with mammalian *tim*. *Current Biology*. **10**(14), 512–513.

Bennett, M.F., Shriner, J., Brown, R.A. (1957) Persistent Tidal Cycles of Spontaneous Motor Activity in the Fiddler Crab, *Uca pugnax*. *Biological Bulletin*. **112**(3), 267.

Bentkowski, P., Markowska, M., Pijanowska, J. (2010) Role of melatonin in the control of depth distribution of *Daphnia magna*. *Hydrobiologia*. **643**(1), 43–50.

Berge, J., Cottier, F., Last, K.S., Varpe, Ø., Leu, E., Søreide, J., Eiane, K., Falk-Petersen, S., Willis, K., Nygård, H., Vogedes, D., Griffiths, C., Johnsen, G., Lorentzen, D., Brierley, A.S. (2009) Diel vertical migration of Arctic zooplankton during the polar night. *Biology*

letters. **5**(1), 69–72.

Bertossa, R.C., van Dijk, J., Diao, W., Saunders, D., Beukeboom, L.W., Beersma, D.G.M. (2013) Circadian Rhythms Differ between Sexes and Closely Related Species of *Nasonia* Wasps. *PLoS ONE*. **8**(3), e60167.

Blau, J., Young, M.W. (1999) Cycling vrille Expression Is Required for a Functional Drosophila Clock. *Cell*. **99**(6), 661–671.

Bolger, A.M., Lohse, M., Usadel, B. (2014) Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*. **30**(15), 2114–2120.

Brady, A.K., Snyder, K.A., Vize, P.D. (2011) Circadian Cycles of Gene Expression in the Coral, *Acropora millepora*. *PLoS ONE*. **6**(9), e25072.

Bregazzi, P.K., Naylor, E. (1972) The Effects of Low Temperature Upon the Locomotor Activity Rhythm of *Talitrus Saltator* (Montagu) (Crustacea: Amphipoda). *Journal of Experimental Biology*. **57**(2), 393.

Brown, F.A., Webb, H.M. (1948) Temperature relations of an endogenous daily rhythmicity in the fiddler crab, *Uca*. *Physiological zoology*. **21**(4), 371–81.

Brown, F., Fingerman, M., Sandeen, M.J., Webb, H.M. (1953) Persistent diurnal and tidal rhythms of colour change in the fiddler crab *Uca pugnax*. *Journal of Experimental Zoology*. **123**, 29–60.

Browne, W.E., Price, A.L., Gerberding, M., Patel, N.H. (2005) Stages of embryonic development in the amphipod crustacean, *Parhyale hawaiensis*. *Genesis*. **42**(3), 124–149.

Busza, A., Emery-Le, M., Rosbash, M., Emery, P. (2004) Roles of the two *Drosophila* CRYPTOCHROME structural domains in circadian photoreception. *Science*. **304**(5676), 1503–6.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L. (2009) BLAST+: architecture and applications. *BMC Bioinformatics*. **10**(1), 421.

Cassone, V.M. (1990) Effects of melatonin on vertebrate circadian systems. *Trends in*

Neurosciences. **13**(11), 457–464.

Cavallari, N., Frigato, E., Vallone, D., Fröhlich, N., Lopez-Olmeda, J.F., Foà, A., Berti, R., Sánchez-Vázquez, F.J., Bertolucci, C., Foulkes, N.S. (2011) A Blind Circadian Clock in Cavefish Reveals that Opsins Mediate Peripheral Clock Photoreception. *PLoS Biology*. **9**(9), e1001142.

Cellier-Michel, S., Berthon, J.L. (2003) Rhythmicity of the Pigments in the Compound Eye of *Daphnia longispina* (Cladocera). *Journal of Freshwater Ecology*. **18**(3), 445–450.

Chabot, C.C., Webb, L.K. (2008) Circadian rhythms of heart rate in freely moving and restrained American lobsters, *Homarus americanus*. *Marine and Freshwater Behaviour and Physiology*. **41**(1), 29–41.

Chang, D.C., McWatters, H.G., Williams, J. a., Gotter, A.L., Levine, J.D., Reppert, S.M. (2003) Constructing a feedback loop with circadian clock molecules from the silkmoth, *Antheraea pernyi. Journal of Biological Chemistry*. **278**(40), 38149–38158.

Chang, D.C., Reppert, S.M. (2003) A Novel C-Terminal Domain of *Drosophila* PERIOD Inhibits dCLOCK:CYCLE-Mediated Transcription. *Current Biology*. **13**(9), 758–762.

Chang, Z., Li, G., Liu, J., Zhang, Y., Ashby, C., Liu, D., Cramer, C.L., Huang, X. (2015) Bridger: a new framework for *de novo* transcriptome assembly using RNA-seq data. *Genome Biology*. **16**(1), 30.

Chaw, R.C., Patel, N.H. (2012) Independent migration of cell populations in the early gastrulation of the amphipod crustacean *Parhyale hawaiensis. Developmental Biology*. **371**(1), 94–109.

Chiu, J.C., Ko, H.W., Edery, I. (2011) NEMO/NLK Phosphorylates PERIOD to Initiate a Time-Delay Phosphorylation Circuit that Sets Circadian Clock Speed. *Cell*. **145**(3), 357–370.

Chiu, J.C., Vanselow, J.T., Kramer, A., Edery, I. (2008) The phospho-occupancy of an atypical SLIMB-binding site on PERIOD that is phosphorylated by DOUBLETIME controls the pace of the clock. *Genes & Development*. **22**(13), 1758–1772.

Christie, A.E., Durkin, C.S., Hartline, N., Ohno, P., Lenz, P.H. (2010a) Bioinformatic analyses of the publicly accessible crustacean expressed sequence tags (ESTs) reveal numerous novel neuropeptide-encoding precursor proteins, including ones from members of several little studied taxa. *General and Comparative Endocrinology*. **167**(1), 164–178.

Christie, A.E., Fontanilla, T.M., Nesbit, K.T., Lenz, P.H. (2013a) Prediction of the protein components of a putative *Calanus finmarchicus* (Crustacea, Copepoda) circadian signaling system using a *de novo* assembled transcriptome. *Comparative biochemistry and physiology. Part D, Genomics & proteomics*. **8**(3), 165–93.

Christie, A.E., McCoole, M.D., Harmon, S.M., Baer, K.N., Lenz, P.H. (2011a) Genomic analyses of the *Daphnia pulex* peptidome. *General and Comparative Endocrinology*. **171**(2), 131–150.

Christie, A.E., Nolan, D.H., Ohno, P., Hartline, N., Lenz, P.H. (2011b) Identification of chelicerate neuropeptides using bioinformatics of publicly accessible expressed sequence tags. *General and Comparative Endocrinology*. **170**(1), 144–155.

Christie, A.E., Roncalli, V., Wu, L.S., Ganote, C.L., Doak, T., Lenz, P.H. (2013b) Peptidergic signaling in *Calanus finmarchicus* (Crustacea, Copepoda): In silico identification of putative peptide hormones and their receptors using a *de novo* assembled transcriptome. *General and Comparative Endocrinology*. **187**, 117–135.

Christie, A.E., Stemmler, E.A., Dickinson, P.S. (2010b) Crustacean neuropeptides. *Cellular and Molecular Life Sciences*. **67**(24), 4135–4169.

Christy, J.H. (1978) Adaptive Significance of Reproductive Cycles in the Fiddler Crab *Uca pugilator*: A Hypothesis. *Science*. **199**(4327), 453–5.

Clark, M.S., Thorne, M. a S., Toullec, J.Y., Meng, Y., Guan, L.L., Peck, L.S., Moore, S. (2011) Antarctic krill 454 pyrosequencing reveals chaperone and stress transcriptome. *PLoS ONE*. **6**(1), 1–17.

Clarke, A., Tyler, P. a. (2008) Adult Antarctic Krill Feeding at Abyssal Depths. *Current Biology*. **18**(4), 282–285.

Conesa, A., Götz, S., García-Gómez, J.M., Terol, J., Talón, M., Robles, M. (2005) Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*. **21**(18), 3674–3676.

Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szcześniak, M.W., Gaffney, D.J., Elo, L.L., Zhang, X., Mortazavi, A. (2016) A survey of best practices for RNA-seq data analysis. *Genome Biology*. **17**(1), 13.

De Coursey, P.J. (1960) Daily Light Sensitivity Rhythm in a Rodent. *Science*. **131**(3392), 33–35.

Cox, M., Kawaguchi, S., King, R., Dholakia, K., Brown, C.T.A. (2015) Internal physiology of live krill revealed using new aquaria techniques and mixed optical microscopy and optical coherence tomography (OCT) imaging techniques. *Marine and Freshwater Behaviour and Physiology*. **48**(6), 455–466.

Cyran, S. a., Buchsbaum, A.M., Reddy, K.L., Lin, M.-C.C., Glossop, N.R.J.J., Hardin, P.E., Young, M.W., Storti, R. V., Blau, J. (2003) *vrille, Pdp1,* and *dClock* form a second feedback loop in the *Drosophila* circadian clock. *Cell*. **112**(3), 329–341.

Daan, S. (2010) A History of Chronobiological Concepts. In U. Albrecht, ed. *The Circadian Clock*. New York, NY: Springer New York.

Daan, S., Pittendrigh, C.S. (1976) A Functional analysis of circadian pacemakers in nocturnal rodents. *Journal of Comparative Physiology A*. **106**(3), 253–266.

Darlington, T.K., Wager-Smith, K., Ceriani, M.F., Staknis, D., Gekakis, N., Steeves, T.D., Weitz, C.J., Takahashi, J.S., Kay, S. a (1998) Closing the circadian loop: CLOCK-induced transcription of its own inhibitors *per* and *tim*. *Science*. **280**(5369), 1599–1603.

Davies, N.J., Krusche, P., Tauber, E., Ott, S. (2015) Analysis of 5' gene regions reveals extraordinary conservation of novel non-coding sequences in a wide range of animals. *BMC Evolutionary Biology*. **15**(1), 227.

Deagle, B.E., Faux, C., Kawaguchi, S., Meyer, B., Jarman, S.N. (2015) Antarctic krill population genomics: apparent panmixia, but genome complexity and large population size muddy the water. *Molecular Ecology*. **24**(19), 4943–4959.

DeBruyne, J.P., Weaver, D.R., Reppert, S.M. (2007) CLOCK and NPAS2 have overlapping roles in the suprachiasmatic circadian clock. *Nature Neuroscience*. **10**(5), 543–545.

Devlin, P.F., Kay, S.A. (1999) Cryptochromes – bringing the blues to circadian rhythms. *Trends in Cell Biology*. **9**(8), 295–298.

Divakaran, O. (1982) Nervous system of *Parhyale hawaiensis* Dana (Crustacea: Amphipoda). *Proceedings of the Indian National Science Academy*. **B48**(2), 218–224.

Dunlap, J., Dunlap, J. (1999) Molecular Bases for Circadian Clocks. *Cell*. **96**(2), 271–290.

Ebihara, S., Marks, T., Hudson, D.J., Menaker, M. (1986) Genetic control of melatonin synthesis in the pineal gland of the mouse. *Science*. **231**(4737), 491–3.

Edgecombe, G.D., Giribet, G., Dunn, C.W., Hejnol, A., Kristensen, R.M., Neves, R.C., Rouse, G.W., Worsaae, K., Sørensen, M. V. (2011) Higher-level metazoan relationships: recent progress and remaining questions. *Organisms Diversity & Evolution*. **11**(2), 151–172.

Emery, P., So, W.V., Kaneko, M., Hall, J.C., Rosbash, M. (1998) CRY, a *Drosophila* Clock and Light-Regulated Cryptochrome, Is a Major Contributor to Circadian Rhythm Resetting and Photosensitivity. *Cell*. **95**(5), 669–679.

Emery, P., Stanewsky, R., Hall, J.C., Rosbash, M. (2000) A unique circadian-rhythm photoreceptor. *Nature*. **404**(6777), 456–7.

Enright, J.T., Hamner, W.M. (1967) Vertical diurnal migration and endogenous rhythmicity. *Science*. **157**(3791), 937–41.

Ettershank, G. (1983) Age structure and cyclical annual size change in the Antarctic krill, *Euphausia superba* dana. *Polar Biology*. **2**(3), 189–193.

Ewer, J., Frisch, B., Hamblen-Coyle, M.J., Rosbash, M., Hall, J.C. (1992) Expression of the *period* clock gene within different cell types in the brain of *Drosophila* adults and mosaic analysis of these cells' influence on circadian behavioral rhythms. *Journal of Neuroscience*. **12**(9), 3321–49.

Extavour, C.G. (2005) The fate of isolated blastomeres with respect to germ cell

formation in the amphipod crustacean *Parhyale hawaiensis*. *Developmental Biology*. **277**(2), 387–402.

Fang, Y., Sathyanarayanan, S., Sehgal, A. (2007) Post-translational regulation of the *Drosophila* circadian clock requires protein phosphatase 1 (PP1). *Genes & Development*. **21**(12), 1506–1518.

Farca Luna, A.J., Hurtado-Zavala, J.I., Reischig, T., Heinrich, R. (2009) Circadian Regulation of Agonistic Behavior in Groups of Parthenogenetic Marbled Crayfish, *Procambarus* sp. *Journal of Biological Rhythms*. **24**(1), 64–72.

Fernández de Miguel, F., Arechiga, H. (1994) Circadian locomotor activity and its entrainment by food in the crayfish *Procambarus clarki*. *Journal of Experimental Biology*. **190**(1), 9–21.

Field, M.D., Maywood, E.S., O'Brien, J.A., Weaver, D.R., Reppert, S.M., Hastings, M.H. (2000) Analysis of Clock Proteins in Mouse SCN Demonstrates Phylogenetic Divergence of the Circadian Clockwork and Resetting Mechanisms. *Neuron*. **25**(2), 437–447.

Fingerman, M. (1970) Circadian rhythm of distal retinal pigment migration in the fiddler crab, *Uca Pugilator* , maintained in constant darkness and its endocrine control. *Journal of Interdisciplinary Cycle Research*. **1**(2), 115–121.

Fingerman, S.W., Fingerman, M. (1977) Circadian variation in the levels of red pigment-dispersing hormone and 5-hydroxytryptamine in the eyestalks of the fiddler crab, *Uca pugilator*. *Comparative Biochemistry and Physiology Part C: Comparative Pharmacology*. **56**(1), 5–8.

Forward, R.B., Bourla, M.H. (2008) Entrainment of the larval release rhythm of the crab *Rhithropanopeus harrisii* (Brachyura: Xanthidae) by cycles in hydrostatic pressure. *Journal of Experimental Marine Biology and Ecology*. **357**(2), 128–133.

Forward, R.B., Lohmann, K.J. (1983) Control of Egg Hatching in the Crab Rhithropanopeus harrisii (Gould). *Biological Bulletin*. **165**(1), 154.

Forward, R.B., Wyatt, L., Clifford, D., Barbour, A. (2007) Endogenous rhythm in activity of an estuarine amphipod, *Talorchestia longicornis*. *Marine and Freshwater Behaviour*

and Physiology. **40**(2), 133–140.

Frazer, T.K., Quetin, L.B., Ross, R.M. (2002) Abundance , sizes and developmental stages of larval krill , *Euphausia superba* , during winter in ice-covered seas west of the Antarctic Peninsula. *Journal of Plankton Research*. **24**(10), 1067–1077.

Frisch, B., Hardin, P.E., Hamblen-Coyle, M.J., Rosbash, M., Hall, J.C. (1994) A promoterless *period* gene mediates behavioral rhythmicity and cyclical per expression in a restricted subset of the *Drosophila* nervous system. *Neuron*. **12**(3), 555–570.

Fu, L., Niu, B., Zhu, Z., Wu, S., Li, W. (2012) CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*. **28**(23), 3150–3152.

Fuentes-Pardo, B., Inclán-Rubio, V. (1987) Caudal photoreceptors synchronize the circadian rhythms in crayfish—I. Synchronization of ERG and locomotor circadian rhythms. *Comparative Biochemistry and Physiology Part A: Physiology*. **86**(3), 523–527.

Fuentes-Pardo, B., Rubio, V.I. (1981) Correlation between motor and electroretinographic circadian rhythms in the crayfish *Procambarus bouvieri* (ortmann). *Comparative Biochemistry and Physiology Part A: Physiology*. **68**(3), 477–485.

Gallego, M., Virshup, D.M. (2007) Post-translational modifications regulate the ticking of the circadian clock. *Nature Reviews Molecular Cell Biology*. **8**(2), 139–148.

Gamble, F.W., Keeble, F.W. (1900) *Hippolyte varians*: a Study in Colour-change. *Quarterly Journal of Microse Science*. **43**, 589–703.

Gard, A.L., Lenz, P.H., Shaw, J.R., Christie, A.E. (2009) Identification of putative peptide paracrines/hormones in the water flea *Daphnia pulex* (Crustacea; Branchiopoda; Cladocera) using transcriptomics and immunohistochemistry. *General and Comparative Endocrinology*. **160**(3), 271–287.

Gaten, E., Tarling, G., Dowse, H., Kyriacou, C., Rosato, E. (2008) Is vertical migration in Antarctic krill (*Euphausia superba*) influenced by an underlying circadian rhythm? *Journal of genetics*. **87**(5), 473–483.

Gaten, E., Wiese, K., Johnson, M.L. (2010) Laboratory-based observations of behaviour in Northern krill (*Meganyctiphanes norvegica* Sars). *Advances in marine biology*. **57**, 231–53.

Gekakis, N., Saez, L., Delahaye-Brown, A.M., Myers, M.P., Sehgal, A., Young, M.W., Weitz, C.J. (1995) Isolation of timeless by PER protein interaction: defective interaction between timeless protein and long-period mutant PERL. *Science*. **270**(5237), 811–5.

Gekakis, N., Staknis, D., Nguyen, H.B., Davis, F.C., Wilsbacher, L.D., King, D.P., Takahashi, J.S., Weitz, C.J. (1998) Role of the CLOCK protein in the mammalian circadian mechanism. *Science*. **280**(5369), 1564–1569.

Gentile, C., Rivas, G.B.S., Meireles-Filho, A.C.A., Lima, J.B.P., Peixoto, A.A. (2009) Circadian Expression of Clock Genes in Two Mosquito Disease Vectors: *cry2* Is Different. *Journal of Biological Rhythms*. **24**(6), 444–451.

Gerber, H.P., Seipel, K., Georgiev, O., Höfferer, M., Hug, M., Rusconi, S., Schaffner, W. (1994) Transcriptional activation modulated by homopolymeric glutamine and proline stretches. *Science*. **263**(5148), 808–811.

Gerberding, M., Browne, W.E., Patel, N.H. (2002) Cell lineage analysis of the amphipod crustacean *Parhyale hawaiensis* reveals an early restriction of cell fates. *Development*. **129**(24), 5789–5801.

Gliwicz, M.Z. (1986) Predation and the evolution of vertical migration in zooplankton. *Nature*. **320**(6064), 746–748.

Glossop, N.R.J., Houl, J.H., Zheng, H., Ng, F.S., Dudek, S.M., Hardin, P.E. (2003) VRILLE Feeds Back to Control Circadian Transcription of Clock in the *Drosophila* Circadian Oscillator. *Neuron*. **37**(2), 249–261.

Godlewska, M., Klusek, Z. (1987) Vertical distribution and diurnal migrations of krill - *Euphausia superba* Dana - from hydroacoustical observations, SIBEX, December 1983/January 1984. *Polar Biology*. **8**(1), 17–22.

Goldstein, J.S., Dubofsky, E.A., Spanier, E. (2015) Into a rhythm: diel activity patterns and behaviour in Mediterranean slipper lobsters, *Scyllarides latus*. *ICES Journal of*

Marine Science: Journal du Conseil. **72**(suppl 1), i147–i154.

Gotter, A.L., Manganaro, T., Weaver, D.R., Kolakowski, L.F., Possidente, B., Sriram, S., MacLaughlin, D.T., Reppert, S.M. (2000) A time-less function for mouse *timeless*. *Nature neuroscience*. **3**(8), 755–6.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D. a, Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*. **29**(7), 644–652.

Granato, F.C., Tironi, T.S., Maciel, F.E., Rosa, C.E., Vargas, M.A., Nery, L.E.M. (2004) Circadian rhythm of pigment migration induced by chromatrophorotropins in melanophores of the crab *Chasmagnathus granulata*. *Comparative Biochemistry and Physiology - A Molecular and Integrative Physiology*. **138**(3), 313–319.

Green, E.W. (2010) *Genetic and bioinformatic screening for behavioural mutations in Drosophila melanogaster*. University of Leicester.

Grotewiel, M.S., Martin, I., Bhandari, P., Cook-Wiens, E. (2005) Functional senescence in *Drosophila melanogaster*. *Ageing Research Reviews*. **4**(3), 372–97.

Gu, H.-F., Xiao, J.-H., Niu, L.-M., Wang, B., Ma, G.-C., Dunn, D.W., Huang, D.-W. (2014) Adaptive evolution of the circadian gene *timeout* in insects. *Scientific Reports*. **4**, 4212.

Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., Macmanes, M.D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C.N., Henschel, R., Leduc, R.D., Friedman, N., Regev, A. (2013) *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*. **8**(8), 1494–512.

Hansen, K.D., Brenner, S.E., Dudoit, S. (2010) Biases in Illumina transcriptome sequencing caused by random hexamer priming. *Nucleic Acids Research*. **38**(12), e131–e131.

Hao, H., Allen, D.L., Hardin, P.E. (1997) A circadian enhancer mediates PER-dependent mRNA cycling in *Drosophila melanogaster*. *Molecular and Cellular Biology*. **17**(7), 3687–93.

Harris, J.E. (1963) The role of endogenous rhythms in vertical migration. *Journal of the Marine Biological Association of the United Kingdom*. **43**(1), 153–166.

Hastings, M.H., Reddy, A.B., Maywood, E.S. (2003) A clockwork web: circadian timing in brain and periphery, in health and disease. *Nature Reviews Neuroscience*. **4**(8), 649–661.

Hayden, P., Lindberg, R.G. (1969) Circadian rhythm in mammalian body temperature entrained by cyclic pressure changes. *Science*. **164**(885), 1288–1289.

Haznedaroglu, B.Z., Reeves, D., Rismani-Yazdi, H., Peccia, J. (2012) Optimization of *de novo* transcriptome assembly from high-throughput short read sequencing data improves functional annotation for non-model organisms. *BMC Bioinformatics*. **13**(1), 170.

Heintzen, C., Liu, Y. (2007) The *Neurospora crassa* Circadian Clock. In *Advances in Genetics*. pp. 25–66.

Helfrich-Förster, C. (2005) Neurobiology of the fruit fly's circadian clock. *Genes, Brain and Behavior*. **4**(2), 65–76.

Helfrich-Förster, C. (1998) Robust circadian rhythmicity of *Drosophila melanogaster* requires the presence of lateral neurons: a brain-behavioral study of disconnected mutants. *Journal of Comparative Physiology A*. **182**(4), 435–453.

Helfrich-Förster, C., Stengl, M., Homberg, U. (1998) Organization of the circadian system in insects. *Chronobiology international*. **15**(6), 567–94.

Helfrich-Förster, C., Winter, C., Hofbauer, A., Hall, J.C., Stanewsky, R. (2001) The circadian clock of fruit flies is blind after elimination of all known photoreceptors. *Neuron*. **30**(1), 249–61.

Hines, M.N. (1954) A Tidal Rhythm in Behavior of Melanophores in Autotomized Legs

of *Uca pugnax*. *Biological Bulletin*. **107**(3), 386–396.

Hirano, Y., Matsuda, T., Kawaguchi, S. (2003) Breeding antarctic krill in captivity. *Marine and Freshwater Behaviour and Physiology*. **36**(4), 259–269.

Hirayama, J., Sassone-Corsi, P. (2005) Structural and functional features of transcription factors controlling the circadian clock. *Current Opinion in Genetics & Development*. **15**(5), 548–556.

Hoadley, K.D., Szmant, A.M., Pyott, S.J. (2011) Circadian Clock Gene Expression in the Coral *Favia fragum* over Diel and Lunar Reproductive Cycles. *PLoS ONE*. **6**(5), e19755.

Hogenesch, J.B., Gu, Y.Z., Jain, S., Bradfield, C.A. (1998) The basic-helix-loop-helix-PAS orphan MOP3 forms transcriptionally active complexes with circadian and hypoxia factors. *Proceedings of the National Academy of Sciences*. **95**(10), 5474–9.

Hooper, L., Harrison, R. a, Summerbell, C.D., Moore, H., Worthington, H. V, Ness, A., Capps, N., Davey Smith, G., Riemersma, R., Ebrahim, S. (2004) Omega 3 fatty acids for prevention and treatment of cardiovascular disease. In L. Hooper, ed. *Cochrane Database of Systematic Reviews*. Chichester, UK: John Wiley & Sons, Ltd, p. CD003177.

van der Horst, G.T., Muijtjens, M., Kobayashi, K., Takano, R., Kanno, S., Takao, M., de Wit, J., Verkerk, A., Eker, A.P., van Leenen, D., Buijs, R., Bootsma, D., Hoeijmakers, J.H., Yasui, A. (1999) Mammalian Cry1 and Cry2 are essential for maintenance of circadian rhythms. *Nature*. **398**(6728), 627–30.

Huang, Z.J., Edery, I., Rosbash, M. (1993) PAS is a dimerization domain common to *Drosophila period* and several transcription factors. *Nature*. **364**(6434), 259–62.

Hung, H.-C., Maurer, C., Kay, S.A., Weber, F. (2007) Circadian transcription depends on limiting amounts of the transcription co-activator nejire/CBP. *Journal of Biological Chemistry*. **282**(43), 31349–57.

Ikeda, T., Dixon, P. (1982) Body shrinkage as a possible over-wintering mechanism of the Antarctic krill, *Euphausia superba* Dana. *Journal of Experimental Marine Biology and Ecology*. **62**(2), 143–151.

Ikeno, T., Tanaka, S.I., Numata, H., Goto, S.G. (2010) Photoperiodic diapause under the control of circadian clock genes in an insect. *BMC biology*. **8**, 116.

Ingram, K.K., Kutowoi, A., Wurm, Y., Shoemaker, D., Meier, R., Bloch, G. (2012) The Molecular Clockwork of the Fire Ant *Solenopsis invicta*. *PLoS ONE*. **7**(11), e45715.

Innes, D.J., Schwartz, S.S., Hebert, P.D.N. (1986) Genotypic diversity and variation in mode of reproduction among populations in the *Daphnia pulex* group. *Heredity*. **57**(3), 345–355.

Jarvis, E. (2014) A day in the life of a *Parhyale* Lab - the Node. *the Node*. [online]. Available from: http://thenode.biologists.com/a-day-in-the-life-of-a-parhyale-lab/lablife/ [Accessed March 21, 2016].

Jeffery, N.W. (2012) The first genome size estimates for six species of krill (Malacostraca, Euphausiidae): Large genomes at the north and south poles. *Polar Biology*. **35**(6), 959–962.

Jennings, B.H. (2011) *Drosophila* – a versatile model in biology & medicine. *Materials Today*. **14**(5), 190–195.

Jin, X., Shearman, L.P., Weaver, D.R., Zylka, M.J., de Vries, G.J., Reppert, S.M. (1999) A molecular mechanism regulating rhythmic output from the suprachiasmatic circadian clock. *Cell*. **96**(1), 57–68.

Johnsen, A., Fidler, A.E., Kuhn, S., Carter, K.L., Hoffmann, A., Barr, I.R., Biard, C., Charmantier, A., Eens, M., Korsten, P., Siitari, H., Tomiuk, J., Kempenaers, B. (2007) Avian Clock gene polymorphism: evidence for a latitudinal cline in allele frequencies. *Molecular Ecology*. **16**(22), 4867–80.

Johnson, W.H. (1938) The Effect of Light on the Vertical Movements of *Acartia clausi* (Giesbrecht). *Biological Bulletin*. **75**(1), 106.

Jones, C.G., Lawton, J.H., Shachak, M. (1994) Organisms as Ecosystem Engineers. *Oikos*. **69**(3), 373.

Jones, D.A. (1970) Population Densities and Breeding in *Eurydice Pulchra* and *Eurydice*

Affinis in Britain. *Journal of the Marine Biological Association of the United Kingdom*. **50**(3), 635.

Jones, D.A., Naylor, E. (1970) The swimming rhythm of the sand beach isopod *Eurydice pulchra*. *Journal of Experimental Marine Biology and Ecology*. **4**(2), 188–199.

Jones, S. (2004) An overview of the basic helix-loop-helix proteins. *Genome Biology*. **5**(6), 226.

Jordan, F. (2009) Keystone species and food webs. *Philosophical Transactions of the Royal Society B: Biological Sciences*. **364**(1524), 1733–1741.

Kallen, J.L., Abrahamse, S.L., Herp, F. Van (1990) Circadian Rhythmicity of the Crustacean Hyperglycemic Hormone (CHH) in the Hemolymph of the Crayfish. *Biological Bulletin*. **179**(3), 351.

Kawaguchi, S., King, R., Meijers, R., Osborn, J.E., Swadling, K.M., Ritz, D. a., Nicol, S. (2010) An experimental aquarium for observing the schooling behaviour of Antarctic krill (*Euphausia superba*). *Deep-Sea Research Part II: Topical Studies in Oceanography*. **57**(7–8), 683–692.

Kawaguchi, S., Yoshida, T., Finley, L., Cramp, P., Nicol, S. (2007) The krill maturity cycle: A conceptual model of the seasonal cycle in Antarctic krill. *Polar Biology*. **30**(6), 689–698.

Kim, E.Y., Ko, H.W., Yu, W., Hardin, P.E., Edery, I. (2007) A DOUBLETIME Kinase Binding Domain on the *Drosophila* PERIOD Protein Is Essential for Its Hyperphosphorylation, Transcriptional Repression, and Circadian Clock Function. *Molecular and Cellular Biology*. **27**(13), 5014–5028.

Kiyohara, Y.B., Tagao, S., Tamanini, F., Morita, A., Sugisawa, Y., Yasuda, M., Yamanaka, I., Ueda, H.R., van der Horst, G.T.J., Kondo, T., Yagita, K. (2006) The BMAL1 C terminus regulates the circadian transcription feedback loop. *Proceedings of the National Academy of Sciences*. **103**(26), 10074–10079.

Kloss, B., Price, J.L., Saez, L., Blau, J., Rothenfluh, A., Wesley, C.S., Young, M.W. (1998) The *Drosophila Clock* Gene double-time Encodes a Protein Closely Related to Human

Casein Kinase Iε. *Cell*. **94**(1), 97–107.

Ko, H.W., Kim, E.Y., Chiu, J., Vanselow, J.T., Kramer, A., Edery, I. (2010) A Hierarchical Phosphorylation Cascade That Regulates the Timing of PERIOD Nuclear Entry Reveals Novel Roles for Proline-Directed Kinases and GSK-3 /SGG in Circadian Clocks. *Journal of Neuroscience*. **30**(38), 12664–12675.

Kobelkova, A., Goto, S.G., Peyton, J.T., Ikeno, T., Lee, R.E., Denlinger, D.L. (2015) Continuous activity and no cycling of clock genes in the Antarctic midge during the polar summer. *Journal of Insect Physiology*. **81**, 90–96.

Koh, K. (2006) JETLAG Resets the *Drosophila* Circadian Clock by Promoting Light-Induced Degradation of TIMELESS. *Science*. **312**(5781), 1809–1812.

Konopka, R.J., Benzer, S. (1971) Clock mutants of *Drosophila melanogaster. Proceedings of the National Academy of Sciences*. **68**(9), 2112–2116.

Konopka, R.J., Pittendrigh, C., Orr, D. (1989) Reciprocal behaviour associated with altered homeostasis and photosensitivity of *Drosophila* clock mutants. *Journal of Neurogenetics*. **6**(1), 1–10.

Konstantinides, N., Averof, M. (2014) A Common Cellular Basis for Muscle Regeneration in Arthropods and Vertebrates. *Science*. **343**(6172), 788–791.

Kontarakis, Z., Pavlopoulos, A. (2014) Transgenesis in Non-model Organisms: The Case of *Parhyale*. In *Methods in Molecular Biology*. pp. 145–181.

Kontarakis, Z., Pavlopoulos, A., Kiupakis, A., Konstantinides, N., Douris, V., Averof, M. (2011) A versatile strategy for gene trapping and trap conversion in emerging model organisms. *Development*. **138**(12), 2625–2630.

Kumar, S., Chen, D., Jang, C., Nall, A., Zheng, X., Sehgal, A. (2014) An ecdysone-responsive nuclear receptor regulates circadian rhythms in *Drosophila*. *Nature Communications*. **5**(May), 5697.

Kume, K., Zylka, M.J., Sriram, S., Shearman, L.P., Weaver, D.R., Jin, X., Maywood, E.S., Hastings, M.H., Reppert, S.M. (1999) mCRY1 and mCRY2 Are Essential Components of

the Negative Limb of the Circadian Clock Feedback Loop. *Cell*. **98**(2), 193–205.

Kyriacou, C.P., Hall, J.C. (1980) Circadian rhythm mutations in *Drosophila melanogaster* affect short-term fluctuations in the male's courtship song. *Proceedings of the National Academy of Sciences*. **77**(11), 6729–6733.

Lakin-Thomas, P.L. (2006) Transcriptional Feedback Oscillators: Maybe, Maybe Not... *Journal of Biological Rhythms*. **21**(2), 83–92.

Lamaze, A., Lamouroux, A., Vias, C., Hung, H.-C., Weber, F., Rouyer, F. (2011) The E3 ubiquitin ligase CTRIP controls CLOCK levels and PERIOD oscillations in *Drosophila*. *EMBO reports*. **12**(6), 549–57.

Lampert, W. (2007) The Adaptive Significance of Diel Vertical Migration of Zooplankton. *Functional Ecology*. **3**(1), 21–27.

Last, K.S., Hobbs, L., Berge, J., Brierley, A.S., Cottier, F. (2016) Moonlight Drives Ocean-Scale Mass Vertical Migration of Zooplankton during the Arctic Winter. *Current Biology*. **26**(2), 244–251.

Lee, C., Bae, K., Edery, I. (1999) PER and TIM inhibit the DNA binding activity of a *Drosophila* CLOCK-CYC/dBMAL1 heterodimer without disrupting formation of the heterodimer: a basis for circadian transcription. *Molecular and Cellular Biology*. **19**(8), 5316–25.

Letunic, I., Doerks, T., Bork, P. (2015) SMART: recent updates, new developments and status in 2015. *Nucleic Acids Research*. **43**(D1), D257–D260.

Li, W., Godzik, A. (2006) Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*. **22**(13), 1658–1659.

Liedvogel, M., Szulkin, M., Knowles, S.C.L., Wood, M.J., Sheldon, B.C. (2009) Phenotypic correlates of *Clock* gene variation in a wild blue tit population: evidence for a role in seasonal timing of reproduction. *Molecular Ecology*. **18**(11), 2444–56.

Lim, C., Lee, J., Choi, C., Kim, J., Doh, E., Choe, J. (2007) Functional Role of CREB-Binding Protein in the Circadian Clock System of *Drosophila melanogaster*. *Molecular and*

Cellular Biology. **27**(13), 4876–4890.

Liubicich, D.M., Serano, J.M., Pavlopoulos, A., Kontarakis, Z., Protas, M.E., Kwan, E., Chatterjee, S., Tran, K.D., Averof, M., Patel, N.H. (2009) Knockdown of *Parhyale* Ultrabithorax recapitulates evolutionary changes in crustacean appendage morphology. *Proceedings of the National Academy of Sciences*. **106**(33), 13892–6.

López-Olmeda, J.F., Madrid, J. a, Sánchez-Vázquez, F.J. (2006) Light and temperature cycles as zeitgebers of zebrafish (*Danio rerio*) circadian activity rhythms. *Chronobiology international*. **23**(3), 537–550.

Ma, M., Bors, E.K., Dickinson, E.S., Kwiatkowski, M. a., Sousa, G.L., Henry, R.P., Smith, C.M., Towle, D.W., Christie, A.E., Li, L. (2009) Characterization of the *Carcinus maenas* neuropeptidome by mass spectrometry and functional genomics. *General and Comparative Endocrinology*. **161**(3), 320–334.

Ma, M., Gard, A.L., Xiang, F., Wang, J., Davoodian, N., Lenz, P.H., Malecha, S.R., Christie, A.E., Li, L. (2010) Combining in silico transcriptome mining and biological mass spectrometry for neuropeptide discovery in the Pacific white shrimp *Litopenaeus vannamei*. *Peptides*. **31**(1), 27–43.

Macquart-Moulin, C. (1999) Diel vertical migration and endogenous swimming rhythm in *Asterope mariae* (Baird) and *Philomedes interpuncta* (Baird) (Crustacea Ostracoda Cypridinidae). *Journal of Plankton Research*. **21**(10), 1891–1910.

Marcus, N.H. (1985) Endogenous control of spawning in a marine copepod. *Journal of Experimental Marine Biology and Ecology*. **91**(3), 263–269.

Marimuthu, G., Rajan, S., Chandrashekaran, M.K. (1981) Social Entrainment of the Circadian-Rhythm in the Flight Activity of the Microchiropteran Bat *Hipposideros-Speoris*. *Behavioral Ecology and Sociobiology*. **8**(2), 147–150.

Martin, A., Serano, J.M., Jarvis, E., Bruce, H.S., Wang, J., Ray, S., Barker, C.A., O'Connell, L.C., Patel, N.H. (2015) CRISPR/Cas9 Mutagenesis Reveals Versatile Roles of Hox Genes in Crustacean Limb Specification and Evolution. *Current Biology*. **26**(1), 14–26.

Martin, J.A., Wang, Z. (2011) Next-generation transcriptome assembly. *Nature Reviews*

Genetics. **12**(10), 671–682.

Martinek, S., Inonog, S., Manoukian, A.S., Young, M.W. (2001) A role for the segment polarity gene shaggy/GSK-3 in the *Drosophila* circadian clock. *Cell*. **105**(6), 769–779.

Martins, M.J.F., Lago-Leston, A., Anjos, A., Duarte, C.M., Agusti, S., Serrão, E.A., Pearson, G.A. (2015) A transcriptome resource for Antarctic krill (*Euphausia superba* Dana) exposed to short-term stress. *Marine genomics*. **23**, 45–7.

Matsumoto, A., Ukai-Tadenuma, M., Yamada, R.G., Houl, J., Uno, K.D., Kasukawa, T., Dauwalder, B., Itoh, T.Q., Takahashi, K., Ueda, R., Hardin, P.E., Tanimura, T., Ueda, H.R. (2007) A functional genomics strategy reveals *clockwork orange* as a transcriptional regulator in the *Drosophila* circadian clock. *Genes & Development*. **21**(13), 1687–1700.

Mazzotta, G.M., De Pittà, C., Benna, C., Tosatto, S.C.E., Lanfranchi, G., Bertolucci, C., Costa, R. (2010) A cry from the krill. *Chronobiology international*. **27**(3), 425–445.

McDonald, M.J., Rosbash, M. (2001) Microarray analysis and organization of circadian gene expression in *Drosophila*. *Cell*. **107**(5), 567–78.

McDonald, M.J., Rosbash, M., Emery, P. (2001) Wild-Type Circadian Rhythmicity Is Dependent on Closely Spaced E Boxes in the *Drosophila timeless* Promoter. *Molecular and Cellular Biology*. **21**(4), 1207–1217.

McNeil, G.P., Zhang, X., Genova, G., Jackson, F.R. (1998) A Molecular Rhythm Mediating Circadian Clock Output in *Drosophila*. *Neuron*. **20**(2), 297–303.

Meissner, R.-A., Kilman, V.L., Lin, J.-M., Allada, R. (2008) TIMELESS Is an Important Mediator of CK2 Effects on Circadian Clock Function In Vivo. *Journal of Neuroscience*. **28**(39), 9732–9740.

Meyer, B., Auerswald, L., Siegel, V., Spahic, C., Pape, C., Fach, B., Teschke, M., Lopata, A., Fuentes, V. (2010) Seasonal variation in body composition, metabolic activity, feeding, and growth of adult krill *Euphausia superba* in the Lazarev Sea. *Marine Ecology Progress Series*. **398**, 1–18.

Michael, T.P., Park, S., Kim, T.-S., Booth, J., Byer, A., Sun, Q., Chory, J., Lee, K. (2007)

Simple Sequence Repeats Provide a Substrate for Phenotypic Variation in the *Neurospora crassa* Circadian Clock. *PLoS ONE*. **2**(8), e795.

Moreno-Sáenz, E., Hernández-Falcón, J., Fuentes-Pardo, B. (1987) Role of the sinus gland in crayfish circadian rhythmicity—II. ERG circadian rhythm. *Comparative Biochemistry and Physiology Part A: Physiology*. **87**(1), 119–125.

Müller, W.E.G., Wang, X., Schröder, H.C., Korzhev, M., Grebenjuk, V.A., Markl, J.S., Jochum, K.P., Pisignano, D., Wiens, M. (2010) A cryptochrome-based photosensory system in the siliceous sponge *Suberites domuncula* (Demospongiae). *FEBS Journal*. **277**(5), 1182–1201.

Murphy, E.., Watkins, J.., Trathan, P.., Reid, K., Meredith, M.., Thorpe, S.., Johnston, N.., Clarke, A., Tarling, G.., Collins, M.., Forcada, J., Shreeve, R.., Atkinson, A., Korb, R., Whitehouse, M.., Ward, P., Rodhouse, P.., Enderlein, P., Hirst, A.., Martin, A.., Hill, S.., Staniland, I.., Pond, D.., Briggs, D.., Cunningham, N.., Fleming, A.. (2007) Spatial and temporal operation of the Scotia Sea ecosystem: a review of large-scale links in a krill centred food web. *Philosophical Transactions of the Royal Society B: Biological Sciences*. **362**(1477), 113–148.

Murphy, E.J., Hofmann, E.E., Watkins, J.L., Johnston, N.M., Piñones, a., Ballerini, T., Hill, S.L., Trathan, P.N., Tarling, G. a., Cavanagh, R. a., Young, E.F., Thorpe, S.E., Fretwell, P. (2013) Comparison of the structure and function of Southern Ocean regional ecosystems: The Antarctic Peninsula and South Georgia. *Journal of Marine Systems*. **109–110**, 22–42.

Myers, A. a. (1985) Shallow-water, coral reef and mangrove Amphipoda (Gammaridea) of Fiji. *Records of the Australian Museum, Supplement*. **5**(December 1985), 1–143.

Myers, E.M. (2003) The circadian control of eclosion. *Chronobiology international*. **20**(5), 775–794.

Myers, M.P., Wager-Smith, K., Rothenfluh-Hilfiker, A., Young, M.W. (1996) Light-Induced Degradation of TIMELESS and Entrainment of the *Drosophila* Circadian Clock. *Science*. **271**(5256), 1736–1740.

Myers, M.P., Wager-Smith, K., Wesley, C.S., Young, M.W., Sehgal, A. (1995) Positional Cloning and Sequence Analysis of the *Drosophila* Clock Gene, *timeless*. *Science*. **270**(5237), 805–808.

Naidoo, N., Song, W., Hunter-Ensor, M., Sehgal, A. (1999) A role for the proteasome in the light response of the timeless clock protein. *Science*. **285**(5434), 1737–41.

Nesbit, K.T., Christie, A.E. (2014) Identification of the molecular components of a *Tigriopus californicus* (Crustacea, Copepoda) circadian clock. *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics*. **12**, 16–44.

Nestorov, P., Battke, F., Levesque, M.P., Gerberding, M. (2013) The Maternal Transcriptome of the Crustacean *Parhyale hawaiensis* Is Inherited Asymmetrically to Invariant Cell Lineages of the Ectoderm and Mesoderm. *PLoS ONE*. **8**(2), e56049.

Newman, S.J., Nicol, S., Ritz, D., Marchant, H. (1999) Susceptibility of Antarctic krill (*Euphausia superba* Dana) to ultraviolet radiation. *Polar Biology*. **22**(1), 50–55.

Newman, S.J., Ritz, D., Nicol, S. (2003) Behavioural reactions of Antarctic krill (*Euphausia superba* Dana) to ultraviolet and photosynthetically active radiation. *Journal of Experimental Marine Biology and Ecology*. **297**(2), 203–217.

Nicol, S. (1990) The Age-Old Problem of Krill Longevity. *BioScience*. **40**(11), 833–836.

Nicol, S., Foster, J., Kawaguchi, S. (2012) The fishery for Antarctic krill - recent developments. *Fish and Fisheries*. **13**(1), 30–40.

Nielsen, J., Peixoto, A.A., Piccin, A., Costa, R., Kyriacou, C.P., Chalmers, D. (1994) Big flies, small repeats: the 'Thr-Gly' region of the *period* gene in Diptera. *Molecular biology and evolution*. **11**(6), 839–53.

O'Malley, K.G., Banks, M.A. (2008) A latitudinal cline in the Chinook salmon (*Oncorhynchus tshawytscha*) *Clock* gene: evidence for selection on PolyQ length variants. *Proceedings of the Royal Society B: Biological Sciences*. **275**(1653), 2813–2821.

van Oort, B.E.H., Tyler, N.J.C., Gerkema, M.P., Folkow, L., Blix, A.S., Stokkan, K.-A.

(2005) Circadian organization in reindeer. *Nature*. **438**(7071), 1095–1096.

Ousley, A., Zafarullah, K., Chen, Y., Emerson, M., Hickman, L., Sehgal, A. (1998) Conserved regions of the *timeless* (*tim*) clock gene in *Drosophila* analyzed through phylogenetic and functional studies. *Genetics*. **148**(2), 815–825.

Ouyang, Y., Andersson, C.R., Kondo, T., Golden, S.S., Johnson, C.H. (1998) Resonating circadian clocks enhance fitness in cyanobacteria. *Proceedings of the National Academy of Sciences*. **95**(15), 8660–4.

Özkaya, Ö., Rosato, E. (2012) The Circadian Clock of the Fly: A Neurogenetics Journey Through Time. In *Advances in Genetics*. pp. 79–123.

Page, T.L., Larimer, J.L. (1972) Entrainment of the circadian locomotor activity rhythm in crayfish. *Journal of Comparative Physiology A*. **78**(2), 107–120.

Page, T.L., Larimer, J.L. (1975) Neural control of circadian rhythmicity in the crayfish. *Journal of Comparative Physiology A*. **97**(1), 59–80.

Palmer, J.D. (1991) Contributions Made to Chronobiology by Studies of Fiddler Crab Rhythms. *Chronobiology International*. **8**(2), 110–130.

Papetti, C., Zane, L., Bortolotto, E., Bucklin, A., Patarnello, T. (2005) Genetic differentiation and local temporal stability of population structure in the euphausiid *Meganyctiphanes norvegica*. *Marine Ecology Progress Series*. **289**, 225–235.

Parchem, R.J., Poulin, F., Stuart, A.B., Amemiya, C.T., Patel, N.H. (2010) BAC library for the amphipod crustacean, *Parhyale hawaiensis*. *Genomics*. **95**(5), 261–7.

Park, J.H., Helfrich-Forster, C., Lee, G., Liu, L., Rosbash, M., Hall, J.C. (2000) Differential regulation of circadian pacemaker output by separate clock genes in *Drosophila*. *Proceedings of the National Academy of Sciences*. **97**(7), 3608–3613.

Pasquali, V. (2015) Locomotor activity rhythms in high arctic freshwater crustacean: *Lepidurus arcticus* (Branchiopoda; Notostraca). *Biological Rhythm Research*. **46**(3), 453–458.

Pavlopoulos, A., Averof, M. (2005) Establishing genetic transformation for comparative

developmental studies in the crustacean *Parhyale hawaiensis*. *Proceedings of the National Academy of Sciences*. **102**(22), 7888–7893.

Pavlopoulos, A., Kontarakis, Z., Liubicich, D.M., Serano, J.M., Akam, M., Patel, N.H., Averof, M. (2009) Probing the evolution of appendage specialization by Hox gene misexpression in an emerging model crustacean. *Proceedings of the National Academy of Sciences*. **106**(33), 13897–902.

Perissinotto, R., Gurney, L., Pakhomov, E. a. (2000) Contribution of heterotrophic material to diet and energy budget of Antarctic krill, *Euphausia superba*. *Marine Biology*. **136**(1), 129–135.

Peschel, N., Chen, K.F., Szabo, G., Stanewsky, R. (2009) Light-Dependent Interactions between the *Drosophila* Circadian Clock Factors Cryptochrome, Jetlag, and Timeless. *Current Biology*. **19**(3), 241–247.

Pevzner, P.A., Tang, H., Waterman, M.S. (2001) An Eulerian path approach to DNA fragment assembly. *Proceedings of the National Academy of Sciences*. **98**(17), 9748–53.

De Pittà, C., Biscontin, A., Albiero, A., Sales, G., Millino, C., Mazzotta, G.M., Bertolucci, C., Costa, R. (2013) The Antarctic Krill *Euphausia superba* Shows Diurnal Cycles of Transcription under Natural Conditions. *PLoS ONE*. **8**(7).

Pittendrigh, C.S. (1960) Circadian rhythms and the circadian organization of living systems. *Cold Spring Harbor Symposia on Quantitative Biology*. **25**, 159–84.

Poovachiranon, S., Boto, K., Duke, N. (1986) Food Preference Studies and Ingestion Rate Measurements of the Mangrove Amphipod *Parhyale hawaiensis* (Dana). *Journal of Experimental Marine Biology and Ecology*. **98**(316), 129–140.

Preitner, N., Damiola, F., Luis-Lopez-Molina, Zakany, J., Duboule, D., Albrecht, U., Schibler, U. (2002) The Orphan Nuclear Receptor REV-ERBα Controls Circadian Transcription within the Positive Limb of the Mammalian Circadian Oscillator. *Cell*. **110**(2), 251–260.

Price, H.J., Boyd, K.R., Boyd, C.M. (1988) Omnivorous feeding behavior of the Antarctic

krill *Euphausia superba*. *Marine Biology*. **97**(1), 67–77.

Price, J.L., Blau, J., Rothenfluh, A., Abodeely, M., Kloss, B., Young, M.W. (1998) *double-time* Is a Novel *Drosophila* Clock Gene that Regulates PERIOD Protein Accumulation. *Cell*. **94**(1), 83–95.

Price, J.L., Dembinska, M.E., Young, M.W., Rosbash, M. (1995) Suppression of PERIOD protein abundance and circadian cycling by the *Drosophila* clock mutation *timeless*. *The EMBO Journal*. **14**(16), 4044–9.

Priyam, A., Woodcroft, B.J., Rai, V., Munagala, A., Moghul, I., Ter, F., Gibbins, M.A., Moon, H., Leonard, G., Rumpf, W., Wurm, Y. (2015) *Sequenceserver: a modern graphical user interface for custom BLAST databases*. Cold Spring Harbor Labs Journals.

Putnam, N.H., Butts, T., Ferrier, D.E.K., Furlong, R.F., Hellsten, U., Kawashima, T., Robinson-Rechavi, M., Shoguchi, E., Terry, A., Yu, J.-K., Benito-Gutiérrez, E.L., Dubchak, I., Garcia-Fernàndez, J., Gibson-Brown, J.J., Grigoriev, I. V, Horton, A.C., de Jong, P.J., Jurka, J., Kapitonov, V. V, Kohara, Y., Kuroki, Y., Lindquist, E., Lucas, S., Osoegawa, K., Pennacchio, L.A., Salamov, A.A., Satou, Y., Sauka-Spengler, T., Schmutz, J., Shin-I, T., Toyoda, A., Bronner-Fraser, M., Fujiyama, A., Holland, L.Z., Holland, P.W.H., Satoh, N., Rokhsar, D.S. (2008) The amphioxus genome and the evolution of the chordate karyotype. *Nature*. **453**(7198), 1064–71.

Quetin, L.B., Ross, R.M. (1984) Depth distribution of developing *Euphausia superba* embryos, predicted from sinking rates. *Marine Biology*. **79**(1), 47–53.

Rao, K.R. (2001) Crustacean Pigmentary-Effector Hormones: Chemistry and Functions of RPCH, PDH, and Related Peptides. *American Zoologist*. **41**(3), 364–379.

Rao, K.V.. (1972) Intertidal amphipods from the Indian coast. *Proceedings of the Indian National Science Academy*. **38**, 190–205.

Rees, D.J., Belzile, C., Glémet, H., Dufresne, F. (2008) Large genomes among caridean shrimp. *Genome*. **51**(2), 159–63.

Rees, D.J., Dufresne, F., Glémet, H., Belzile, C. (2007) Amphipod genome sizes: first estimates for Arctic species reveal genomic giants. *Genome*. **50**(2), 151–8.

Regier, J.C., Shultz, J.W., Zwick, A., Hussey, A., Ball, B., Wetzer, R., Martin, J.W., Cunningham, C.W. (2010) Arthropod relationships revealed by phylogenomic analysis of nuclear protein-coding sequences. *Nature*. **463**(7284), 1079–1083.

Rehm, E.J., Hannibal, R.L., Chaw, R.C., Vargas-Vila, M. a., Patel, N.H. (2009a) Antibody Staining of *Parhyale hawaiensis* Embryos. *Cold Spring Harbor Protocols*. **2009**(1), pdb.prot5129-prot5129.

Rehm, E.J., Hannibal, R.L., Chaw, R.C., Vargas-Vila, M. a., Patel, N.H. (2009b) In Situ Hybridization of Labeled RNA Probes to Fixed *Parhyale hawaiensis* Embryos. *Cold Spring Harbor Protocols*. **2009**(1), pdb.prot5130-prot5130.

Rehm, E.J., Hannibal, R.L., Chaw, R.C., Vargas-Vila, M. a., Patel, N.H. (2009c) The Crustacean *Parhyale hawaiensis*: A New Model for Arthropod Development. *Cold Spring Harbor Protocols*. **2009**(1), pdb.emo114-emo114.

Reid, K., Croxall, J.P. (2001) Environmental response of upper trophic-level predators reveals a system change in an Antarctic marine ecosystem. *Proceedings of the Royal Society B: Biological Sciences*. **268**(1465), 377–384.

Reitzel, A.M., Behrendt, L., Tarrant, A.M. (2010) Light Entrained Rhythmic Gene Expression in the Sea Anemone *Nematostella vectensis*: The Evolution of the Animal Circadian Clock. *PLoS ONE*. **5**(9), e12805.

Renn, S.C.., Park, J.H., Rosbash, M., Hall, J.C., Taghert, P.H. (1999) A *pdf* Neuropeptide Gene Mutation and Ablation of PDF Neurons Each Cause Severe Abnormalities of Behavioral Circadian Rhythms in *Drosophila*. *Cell*. **99**(7), 791–802.

Reppert, S.M. (2007) The ancestral circadian clock of monarch butterflies: Role in time-compensated sun compass orientation. *Cold Spring Harbor Symposia on Quantitative Biology*. **72**, 113–118.

Reppert, S.M., Weaver, D.R. (2000) Comparing Clockworks: Mouse versus Fly. *Journal of Biological Rhythms*. **15**(5), 357–364.

Richier, B., Michard-Vanhee, C., Lamouroux, a., Papin, C., Rouyer, F. (2008) The Clockwork Orange *Drosophila* Protein Functions as Both an Activator and a Repressor

of Clock Gene Expression. *Journal of Biological Rhythms*. **23**(2), 103–116.

Riihimaa, A., Kimura, M. (1988) A mutant strain of *Chymomyza costata* (Diptera: Drosophilidae) insensitive to diapause-inducing action of photoperiod. *Physiological Entomology*. **13**(4), 441–445.

Ritz, D.A., Cromer, L., Swadling, K.M., Nicol, S., Osborn, J. (2003) Heart rate as a measure of stress in Antarctic krill, *Euphausia superba*. *Journal of the Marine Biological Association of the United Kingdom*. **83**(2), 329–330.

Rivera, A.S., Ozturk, N., Fahey, B., Plachetzki, D.C., Degnan, B.M., Sancar, A., Oakley, T.H. (2012) Blue-light-receptive cryptochrome is expressed in a sponge eye lacking neurons and opsin. *Journal of Experimental Biology*. **215**(Pt 8), 1278–86.

Robertson, G., Schein, J., Chiu, R., Corbett, R., Field, M., Jackman, S.D., Mungall, K., Lee, S., Okada, H.M., Qian, J.Q., Griffith, M., Raymond, A., Thiessen, N., Cezard, T., Butterfield, Y.S., Newsome, R., Chan, S.K., She, R., Varhol, R., Kamoh, B., Prabhu, A.-L., Tam, A., Zhao, Y., Moore, R.A., Hirst, M., Marra, M.A., Jones, S.J.M., Hoodless, P.A., Birol, I. (2010) *De novo* assembly and analysis of RNA-seq data. *Nature Methods*. **7**(11), 909–12.

Robinson, M.D., Oshlack, A. (2010) A scaling normalization method for differential expression analysis of RNA-seq data. *Genome biology*. **11**(3), R25.

Rodriguez, E. (1996) Light responses and eye regression in cavemicolous animals from the Jameos del Agua (Lanzarote, Canary Islands). In J. Ott, M. Stachowitsch, & F. Uiblein, eds. *Deep-sea and extreme shallow-water habitats: affinities and adaptations.* Austrian Academy of Sciences, Vienna, pp. 145–150.

Rodríguez-Sosa, L., Calderón-Rosete, G., Flores, G. (2008) Circadian and ultradian rhythms in the crayfish caudal photoreceptor. *Synapse*. **62**(9), 643–652.

Romanowski, A., Garavaglia, M.J., Goya, M.E., Ghiringhelli, P.D., Golombek, D.A. (2014) Potential Conservation of Circadian Clock Proteins in the phylum Nematoda as Revealed by Bioinformatic Searches. *PLoS ONE*. **9**(11), e112871.

Rosato, E., Codd, V., Mazzotta, G., Piccin, A., Zordan, M., Costa, R., Kyriacou, C.P.

(2001) Light-dependent interaction between *Drosophila* CRY and the clock protein PER mediated by the carboxy terminus of CRY. *Current Biology*. **11**(12), 909–917.

Rosato, E., Kyriacou, C.P. (2006) Analysis of locomotor activity rhythms in *Drosophila*. *Nature Protocols*. **1**(2), 559–568.

Rosato, E., Kyriacou, C.P. (2011) The role of natural selection in circadian behaviour: a molecular-genetic approach. *Essays In Biochemistry*. **49**(1), 71–85.

Rosato, E., Piccin, A., Kyriacou, C.P. (1997) Circadian rhythms: From behaviour to molecules. *BioEssays*. **19**(12), 1075–1082.

Ross, R.M., Quetin, L.B. (1986) How productive are Antarctic krill ? *BioScience*. **36**(4), 264–269.

Ross, R.M., Quetin, L.B. (1983) Spawning frequency and fecundity of the Antarctic krill *Euphausia superba*. *Marine Biology*. **77**(3), 201–205.

Rubin, E.B., Shemesh, Y., Cohen, M., Elgavish, S., Robertson, H.M., Bloch, G. (2006) Molecular and phylogenetic analyses reveal mammalian-like clockwork in the honey bee (*Apis mellifera*) and shed new light on the molecular evolution of the circadian clock. *Genome Research*. **16**(11), 1352–1365.

Rutila, J.E., Suri, V., Le, M., So, W.V., Rosbash, M., Hall, J.C. (1998) Cycle is a second bHLH-PAS clock protein essential for circadian rhythmicity and transcription of *Drosophila period* and *timeless*. *Cell*. **93**(5), 805–814.

Saez, L., Young, M.W. (1996) Regulation of nuclear entry of the *Drosophila* clock proteins period and timeless. *Neuron*. **17**(5), 911–920.

Saigusa, M. (1980) Entrainment of a semilunar rhythm by a simulated moonlight cycle in the terrestrial crab, *Sesarma haematocheir*. *Oecologia*. **46**(1), 38–44.

Sakai, T., Ishida, N. (2001) Circadian rhythms of female mating activity governed by clock genes in *Drosophila*. *Proceedings of the National Academy of Sciences*. **98**(16), 9221–5.

Saleem, Q., Anand, a., Jain, S., Brahmachari, S.K. (2001) The polyglutamine motif is

highly conserved at the *Clock* locus in various organisms and is not polymorphic in humans. *Human Genetics*. **109**(2), 136–142.

Sandrelli, F., Costa, R., Kyriacou, C.P., Rosato, E. (2008) Comparative analysis of circadian clock genes in insects. *Insect Molecular Biology*. **17**(5), 447–463.

Sandrelli, F., Tauber, E., Pegoraro, M., Mazzotta, G., Cisotto, P., Landskron, J., Stanewsky, R., Piccin, A., Rosato, E., Zordan, M., Costa, R., Kyriacou, C.P. (2007) A molecular basis for natural selection at the *timeless* locus in *Drosophila melanogaster*. *Science*. **316**(5833), 1898–1900.

Sanggaard, K.W., Bechsgaard, J.S., Fang, X., Duan, J., Dyrlund, T.F., Gupta, V., Jiang, X., Cheng, L., Fan, D., Feng, Y., Han, L., Huang, Z., Wu, Z., Liao, L., Settepani, V., Thøgersen, I.B., Vanthournout, B., Wang, T., Zhu, Y., Funch, P., Enghild, J.J., Schauser, L., Andersen, S.U., Villesen, P., Schierup, M.H., Bilde, T., Wang, J. (2014) Spider genomes provide insight into composition and evolution of venom and silk. *Nature Communications*. **5**, 3765.

Sarov-Blat, L., So, W.V. V, Liu, L., Rosbash, M. (2000) The *Drosophila takeout* gene is a novel molecular link between circadian rhythms and feeding behavior. *Cell*. **101**(6), 647–656.

Sathyanarayanan, S., Zheng, X., Xiao, R., Sehgal, A. (2004) Posttranslational Regulation of *Drosophila* PERIOD Protein by Protein Phosphatase 2A. *Cell*. **116**(4), 603–615.

Sato, T.K., Panda, S., Miraglia, L.J., Reyes, T.M., Rudic, R.D., McNamara, P., Naik, K.A., FitzGerald, G.A., Kay, S.A., Hogenesch, J.B. (2004) A functional genomics strategy reveals Rora as a component of the mammalian circadian clock. *Neuron*. **43**(4), 527–37.

Sawyer, L.A., Hennessy, J.M., Peixoto, A.A., Rosato, E., Parkinson, H., Costa, R., Kyriacou, C.P. (1997) Natural variation in a *Drosophila* clock gene and temperature compensation. *Science*. **278**(5346), 2117–20.

Sbragaglia, V., Lamanna, F., M. Mat, A., Rotllant, G., Joly, S., Ketmaier, V., de la Iglesia, H.O., Aguzzi, J. (2015) Identification, Characterization, and Diel Pattern of Expression of

Canonical Clock Genes in *Nephrops norvegicus* (Crustacea: Decapoda) Eyestalk. *PLOS ONE*. **10**(11), e0141893.

Scapini, F., Rossano, C., Marchetti, G.M., Morgan, E. (2005) The role of the biological clock in the sun compass orientation of free-running individuals of *Talitrus saltator*. *Animal Behaviour*. **69**(4), 835–843.

Schmidt, K., Atkinson, A., Steigenberger, S., Fielding, S., Lindsay, M.C.M., Pond, D.W., Tarling, G. a., Klevjer, T. a., Allen, C.S., Nicol, S., Achterberg, E.P. (2011) Seabed foraging by Antarctic krill: Implications for stock assessment, bentho-pelagic coupling, and the vertical transfer of iron. . **56**(4), 1411–1428.

Schultz, J., Milpetz, F., Bork, P., Ponting, C.P. (1998) SMART, a simple modular architecture research tool: Identification of signaling domains. *Proceedings of the National Academy of Sciences*. **95**(11), 5857–5864.

Schwarzenberger, A., Wacker, A. (2015) Melatonin synthesis follows a daily cycle in *Daphnia. Journal of Plankton Research*. **37**(3), 636–644.

Seear, P., Tarling, G. a., Teschke, M., Meyer, B., Thorne, M. a S., Clark, M.S., Gaten, E., Rosato, E. (2009) Effects of simulated light regimes on gene expression in Antarctic krill (*Euphausia superba* Dana). *Journal of Experimental Marine Biology and Ecology*. **381**(1), 57–64.

Sehadová, H., Shao, Q.-M., Sehnal, F., Takeda, M. (2007) Neurohormones as putative circadian clock output signals in the central nervous system of two cricket species. *Cell and Tissue Research*. **328**(1), 239–255.

Sehgal, A., Price, J.L., Man, B., Young, M.W. (1994) Loss of circadian behavioral rhythms and *per* RNA oscillations in the *Drosophila* mutant *timeless*. *Science*. **263**(5153), 1603–1606.

Serano, J.M., Martin, A., Liubicich, D.M., Jarvis, E., Bruce, H.S., La, K., Browne, W.E., Grimwood, J., Patel, N.H. (2016) Comprehensive analysis of Hox gene expression in the amphipod crustacean *Parhyale hawaiensis*. *Developmental Biology*. **409**(1), 297–309.

Shafer, O.T., Rosbash, M., Truman, J.W. (2002) Sequential nuclear accumulation of the

clock proteins period and timeless in the pacemaker neurons of *Drosophila melanogaster*. *Journal of Neuroscience*. **22**(14), 5946–54.

Shin, S.W., Zou, Z., Saha, T.T., Raikhel, A.S. (2012) bHLH-PAS heterodimer of methoprene-tolerant and Cycle mediates circadian expression of juvenile hormone-induced mosquito genes. *Proceedings of the National Academy of Sciences*. **109**(41), 16576–81.

Shoemaker, C.R. (1956) Observations on the Amphipod Genus *Parhyale*. *Proceedings of the United States National Museum*. **106**(3372), 345–358.

Shoguchi, E., Tanaka, M., Shinzato, C., Kawashima, T., Satoh, N. (2013) A genome-wide survey of photoreceptor and circadian genes in the coral, *Acropora digitifera*. *Gene*. **515**(2), 426–31.

Simionato, E., Ledent, V., Richards, G., Thomas-Chollier, M., Kerner, P., Coornaert, D., Degnan, B.M., Vervoort, M. (2007) Origin and diversification of the basic helix-loop-helix gene family in metazoans: insights from comparative genomics. *BMC Evolutionary Biology*. **7**(1), 33.

Skopik, S.D., Pittendrigh, C.S. (1967) Circadian systems, II. The oscillation in the individual *Drosophila* pupa; its independence of developmental stage. *Proceedings of the National Academy of Sciences*. **58**(5), 1862–9.

Skov, M.W., Hartnoll, R.G., Ruwa, R.K., Shunula, J.P., Vannini, M., Cannicci, S. (2005) Marching to a different drummer: Crabs synchronize reproduction to a 14-month lunar-tidal cycle. *Ecology*. **86**(5), 1164–1171.

Smith-Unna, R.D., Boursnell, C., Patro, R., Hibberd, J.M., Kelly, S. (2015) TransRate: reference free quality assessment of *de-novo* transcriptome assemblies. *bioRxiv*, 21626. [online]. Available from: http://biorxiv.org/content/early/2015/06/27/021626.abstract [Accessed June 29, 2015].

Sosa, L.R., de la Vega, M.T., Aréchiga, H. (1994) Circadian rhythm of content of red pigment-concentrating hormone in the crayfish eyestalk. *Comparative Biochemistry*

and Physiology Part C: Pharmacology, Toxicology and Endocrinology. **109**(1), 101–108.

Strauss, J., Dircksen, H. (2010) Circadian clocks in crustaceans: identified neuronal and cellular systems. *Frontiers in Bioscience*. **15**, 1040–74.

Surget-Groba, Y., Montoya-Burgos, J.I. (2010) Optimization of *de novo* transcriptome assembly from next-generation sequencing data. *Genome Research*. **20**(10), 1432–1440.

Takahata, S., Ozaki, T., Mimura, J., Kikuchi, Y., Sogawa, K., Fujii-Kuriyama, Y. (2000) Transactivation mechanisms of mouse clock transcription factors, mClock and mArnt3. *Genes to Cells*. **5**(9), 739–747.

Taki, K., Hayashi, T., Naganobu, M. (2005) Characteristics of seasonal variation in diurnal vertical migration and aggregation of Antarctic krill (*Euphausia superba*) in the Scotia Sea, using Japanese fishery data. *CCAMLR Science*. **12**, 163–172.

Tarling, G.A., Ensor, N.S., Fregin, T., Goodall-Copestake, W.P., Fretwell, P. (2010) An Introduction to the Biology of Northern Krill (*Meganyctiphanes norvegica* Sars). In *Advances in Marine Biology*. pp. 1–40.

Tarling, G. a., Johnson, M.L. (2006) Satiation gives krill that sinking feeling. *Current Biology*. **16**(3), R83–R84.

Tarling, G. a., Klevjer, T., Fielding, S., Watkins, J., Atkinson, A., Murphy, E., Korb, R., Whitehouse, M., Leaper, R. (2009) Variability and predictability of Antarctic krill swarm structure. *Deep-Sea Research Part I: Oceanographic Research Papers*. **56**(11), 1994–2012.

Tauber, E., Roe, H., Costa, R., Hennessy, J.M., Kyriacou, C.P. (2003) Temporal Mating Isolation Driven by a Behavioral Gene in *Drosophila*. *Current Biology*. **13**(2), 140–145.

Terrapon, N., Li, C., Robertson, H.M., Ji, L., Meng, X., Booth, W., Chen, Z., Childers, C.P., Glastad, K.M., Gokhale, K., Gowin, J., Gronenberg, W., Hermansen, R.A., Hu, H., Hunt, B.G., Huylmans, A.K., Khalil, S.M.S., Mitchell, R.D., Munoz-Torres, M.C., Mustard, J.A., Pan, H., Reese, J.T., Scharf, M.E., Sun, F., Vogel, H., Xiao, J., Yang, W., Yang, Z., Yang, Z., Zhou, J., Zhu, J., Brent, C.S., Elsik, C.G., Goodisman, M.A.D., Liberles, D.A., Roe, R.M.,

Vargo, E.L., Vilcinskas, A., Wang, J., Bornberg-Bauer, E., Korb, J., Zhang, G., Liebig, J. (2014) Molecular traces of alternative social organization in a termite genome. *Nature communications*. **5**, 3636.

Teschke, M., Wendt, S., Kawaguchi, S., Kramer, A., Meyer, B. (2011) A Circadian Clock in Antarctic Krill: An Endogenous Timing System Governs Metabolic Output Rhythms in the Euphausid Species *Euphausia superba*. *PLoS ONE*. **6**(10), e26090.

Thomas, P.G., Ikeda, T. (1987) Sexual regression, shrinkage, re-maturation and growth of spent female *Euphausia superba* in the laboratory. *Marine Biology*. **95**(3), 357–363.

Tilden, A.R., McCoole, M.D., Harmon, S.M., Baer, K.N., Christie, A.E. (2011) Genomic identification of a putative circadian system in the cladoceran crustacean *Daphnia pulex*. *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics*. **6**(3), 282–309.

Tomioka, K., Matsumoto, A. (2015) Circadian molecular clockworks in non-model insects. *Current Opinion in Insect Science*. **7**, 58–64.

Tosches, M.A., Bucher, D., Vopalensky, P., Arendt, D. (2014) Melatonin signaling controls circadian swimming behavior in marine zooplankton. *Cell*. **159**(1), 46–57.

Toullec, J.-Y., Corre, E., Bernay, B., Thorne, M.A.S., Cascella, K., Ollivaux, C., Henry, J., Clark, M.S. (2013) Transcriptome and Peptidome Characterisation of the Main Neuropeptides and Peptidic Hormones of a Euphausiid: The Ice Krill, *Euphausia crystallorophias*. *PLoS ONE*. **8**(8), e71609.

Verde, M. a., Barriga-Montoya, C., Fuentes-Pardo, B. (2007) Pigment dispersing hormone generates a circadian response to light in the crayfish, *Procambarus clarkii*. *Comparative biochemistry and physiology. Part A, Molecular & integrative physiology*. **147**(4), 983–92.

Vitaterna, M.H., King, D.P., Chang, A.M., Kornhauser, J.M., Lowrey, P.L., McDonald, J.D., Dove, W.F., Pinto, L.H., Turek, F.W., Takahashi, J.S. (1994) Mutagenesis and mapping of a mouse gene, *Clock*, essential for circadian behavior. *Science*. **264**(5159), 719–25.

Waddy, S.L., Aiken, D.E. (1999) Timing of the metamorphic molt of the American lobster (*Homarus americanus*) is governed by a population-based, photoperiodically entrained daily rhythm. *Canadian Journal of Fisheries and Aquatic Sciences*. **56**(12), 2324–2330.

Wagner, G.P., Kin, K., Lynch, V.J. (2012) Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory in Biosciences*. **131**(4), 281–285.

Wang, Z., Gerstein, M., Snyder, M. (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*. **10**(1), 57–63.

Warman, C.G., Naylor, E. (1995) Evidence for multiple, cue-specific circatidal clocks in the shore crab *Carcinus maenas*. *Journal of Experimental Marine Biology and Ecology*. **189**(1–2), 93–101.

Webb, H.M. (1950) Diurnal variations of response to light in the fiddler crab, Uca. *Physiological Zoology*. **23**(4), 316–37.

Werren, J.H., Richards, S., Desjardins, C.A., Niehuis, O., Gadau, J., Colbourne, J.K., Beukeboom, L.W., Desplan, C., Elsik, C.G., Grimmelikhuijzen, C.J.P., Kitts, P., Lynch, J.A., Murphy, T., Oliveira, D.C.S.G., Smith, C.D., van de Zande, L., Worley, K.C., Zdobnov, E.M., Aerts, M., Albert, S., Anaya, V.H., Anzola, J.M., Barchuk, A.R., Behura, S.K., Bera, A.N., Berenbaum, M.R., Bertossa, R.C., Bitondi, M.M.G., Bordenstein, S.R., Bork, P., Bornberg-Bauer, E., Brunain, M., Cazzamali, G., Chaboub, L., Chacko, J., Chavez, D., Childers, C.P., Choi, J.-H., Clark, M.E., Claudianos, C., Clinton, R.A., Cree, A.G., Cristino, A.S., Dang, P.M., Darby, A.C., de Graaf, D.C., Devreese, B., Dinh, H.H., Edwards, R., Elango, N., Elhaik, E., Ermolaeva, O., Evans, J.D., Foret, S., Fowler, G.R., Gerlach, D., Gibson, J.D., Gilbert, D.G., Graur, D., Gründer, S., Hagen, D.E., Han, Y., Hauser, F., Hultmark, D., Hunter, H.C., Hurst, G.D.D., Jhangian, S.N., Jiang, H., Johnson, R.M., Jones, A.K., Junier, T., Kadowaki, T., Kamping, A., Kapustin, Y., Kechavarzi, B., Kim, J., Kim, J., Kiryutin, B., Koevoets, T., Kovar, C.L., Kriventseva, E. V, Kucharski, R., Lee, H., Lee, S.L., Lees, K., Lewis, L.R., Loehlin, D.W., Logsdon, J.M., Lopez, J.A., Lozado, R.J., Maglott, D., Maleszka, R., Mayampurath, A., Mazur, D.J., McClure, M.A., Moore, A.D.,

Waddy, S.L., Aiken, D.E. (1999) Timing of the metamorphic molt of the American lobster (*Homarus americanus*) is governed by a population-based, photoperiodically entrained daily rhythm. *Canadian Journal of Fisheries and Aquatic Sciences*. **56**(12), 2324–2330.

Wagner, G.P., Kin, K., Lynch, V.J. (2012) Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory in Biosciences*. **131**(4), 281–285.

Wang, Z., Gerstein, M., Snyder, M. (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*. **10**(1), 57–63.

Warman, C.G., Naylor, E. (1995) Evidence for multiple, cue-specific circatidal clocks in the shore crab *Carcinus maenas*. *Journal of Experimental Marine Biology and Ecology*. **189**(1–2), 93–101.

Webb, H.M. (1950) Diurnal variations of response to light in the fiddler crab, Uca. *Physiological Zoology*. **23**(4), 316–37.

Werren, J.H., Richards, S., Desjardins, C.A., Niehuis, O., Gadau, J., Colbourne, J.K., Beukeboom, L.W., Desplan, C., Elsik, C.G., Grimmelikhuijzen, C.J.P., Kitts, P., Lynch, J.A., Murphy, T., Oliveira, D.C.S.G., Smith, C.D., van de Zande, L., Worley, K.C., Zdobnov, E.M., Aerts, M., Albert, S., Anaya, V.H., Anzola, J.M., Barchuk, A.R., Behura, S.K., Bera, A.N., Berenbaum, M.R., Bertossa, R.C., Bitondi, M.M.G., Bordenstein, S.R., Bork, P., Bornberg-Bauer, E., Brunain, M., Cazzamali, G., Chaboub, L., Chacko, J., Chavez, D., Childers, C.P., Choi, J.-H., Clark, M.E., Claudianos, C., Clinton, R.A., Cree, A.G., Cristino, A.S., Dang, P.M., Darby, A.C., de Graaf, D.C., Devreese, B., Dinh, H.H., Edwards, R., Elango, N., Elhaik, E., Ermolaeva, O., Evans, J.D., Foret, S., Fowler, G.R., Gerlach, D., Gibson, J.D., Gilbert, D.G., Graur, D., Gründer, S., Hagen, D.E., Han, Y., Hauser, F., Hultmark, D., Hunter, H.C., Hurst, G.D.D., Jhangian, S.N., Jiang, H., Johnson, R.M., Jones, A.K., Junier, T., Kadowaki, T., Kamping, A., Kapustin, Y., Kechavarzi, B., Kim, J., Kim, J., Kiryutin, B., Koevoets, T., Kovar, C.L., Kriventseva, E. V, Kucharski, R., Lee, H., Lee, S.L., Lees, K., Lewis, L.R., Loehlin, D.W., Logsdon, J.M., Lopez, J.A., Lozado, R.J., Maglott, D., Maleszka, R., Mayampurath, A., Mazur, D.J., McClure, M.A., Moore, A.D.,

Morgan, M.B., Muller, J., Munoz-Torres, M.C., Muzny, D.M., Nazareth, L. V, Neupert, S., Nguyen, N.B., Nunes, F.M.F., Oakeshott, J.G., Okwuonu, G.O., Pannebakker, B.A., Pejaver, V.R., Peng, Z., Pratt, S.C., Predel, R., Pu, L.-L., Ranson, H., Raychoudhury, R., Rechtsteiner, A., Reese, J.T., Reid, J.G., Riddle, M., Robertson, H.M., Romero-Severson, J., Rosenberg, M., Sackton, T.B., Sattelle, D.B., Schlüns, H., Schmitt, T., Schneider, M., Schüler, A., Schurko, A.M., Shuker, D.M., Simões, Z.L.P., Sinha, S., Smith, Z., Solovyev, V., Souvorov, A., Springauf, A., Stafflinger, E., Stage, D.E., Stanke, M., Tanaka, Y., Telschow, A., Trent, C., Vattathil, S., Verhulst, E.C., Viljakainen, L., Wanner, K.W., Waterhouse, R.M., Whitfield, J.B., Wilkes, T.E., Williamson, M., Willis, J.H., Wolschin, F., Wyder, S., Yamada, T., Yi, S. V, Zecher, C.N., Zhang, L., Gibbs, R.A. (2010) Functional and evolutionary insights from the genomes of three parasitoid *Nasonia* species. *Science*. **327**(5963), 343–8.

Whitehouse, M.J., Atkinson, A., Rees, A.P. (2011) Close coupling between ammonium uptake by phytoplankton and excretion by Antarctic krill, *Euphausia superba*. *Deep Sea Research Part I: Oceanographic Research Papers*. **58**(7), 725–732.

Wilcockson, D.C., Zhang, L., Hastings, M.H., Kyriacou, C.P., Webster, S.G. (2011) A novel form of pigment-dispersing hormone in the central nervous system of the intertidal marine isopod, *Eurydice pulchra* (leach). *Journal of Comparative Neurology*. **519**(3), 562–575.

Xie, Y., Wu, G., Tang, J., Luo, R., Patterson, J., Liu, S., Huang, W., He, G., Gu, S., Li, S., Zhou, X., Lam, T.-W., Li, Y., Xu, X., Wong, G.K.-S., Wang, J. (2014) SOAPdenovo-Trans: *de novo* transcriptome assembly with short RNA-Seq reads. *Bioinformatics*. **30**(12), 1660–1666.

Yang, J.S., Dai, Z.M., Yang, F., Yang, W.J. (2006) Molecular cloning of Clock cDNA from the prawn, *Macrobrachium rosenbergii*. *Brain Research*. **1067**(1), 13–24.

Yang, Z., Sehgal, A. (2001) Role of Molecular Oscillations in Generating Behavioral Rhythms in *Drosophila*. *Neuron*. **29**(2), 453–467.

Yanovsky, M.J., Kay, S. a (2003) Living by the calendar: how plants know when to flower. *Nature Reviews Molecular Cell Biology*. **4**(4), 265–276.

Yu, W., Houl, J.H., Hardin, P.E. (2011) NEMO Kinase Contributes to Core Period Determination by Slowing the Pace of the *Drosophila* Circadian Oscillator. *Current Biology*. **21**(9), 756–761.

Yuan, Q., Metterville, D., Briscoe, A.D., Reppert, S.M. (2007) Insect cryptochromes: Gene duplication and loss define diverse ways to construct insect circadian clocks. *Molecular Biology and Evolution*. **24**(4), 948–955.

Zane, L., Ostellari, L., Maccatrozzo, L., Bargelloni, L., Cuzin-Roudy, J., Buchholz, F., Patarnello, T. (2000) Genetic differentiation in a pelagic crustacean (*Meganyctiphanes norvegica* : Euphausiacea) from the North East Atlantic and the Mediterranean Sea. *Marine Biology*. **136**(2), 191–199.

Zantke, J., Ishikawa-Fujiwara, T., Arboleda, E., Lohs, C., Schipany, K., Hallay, N., Straw, A.D., Todo, T., Tessmar-Raible, K. (2013) Circadian and Circalunar Clock Interactions in a Marine Annelid. *Cell Reports*. **5**(1), 99–113.

Zeng, H., Qian, Z., Myers, M.P., Rosbash, M. (1996) A light-entrainment mechanism for the *Drosophila* circadian clock. *Nature*. **380**(6570), 129–35.

Zhang, L., Hastings, M.H., Green, E.W., Tauber, E., Sladek, M., Webster, S.G., Kyriacou, C.P., Wilcockson, D.C. (2013) Dissociation of circadian and circatidal timekeeping in the marine crustacean *Eurydice pulchra*. *Current biology*. **23**(19), 1863–73.

Zheng, X., Koh, K., Sowcik, M., Smith, C.J., Chen, D., Wu, M.N., Sehgal, A. (2009) An Isoform-Specific Mutant Reveals a Role of PDP1 in the Circadian Oscillator. *Journal of Neuroscience*. **29**(35), 10920–10927.

Zhu, H., Sauman, I., Yuan, Q., Casselman, A., Emery-Le, M., Emery, P., Reppert, S.M. (2008) Cryptochromes define a novel circadian clock mechanism in monarch butterflies that may underlie sun compass navigation. *PLoS Biology*. **6**(1), 0138–0155.

Zhu, H., Yuan, Q., Froy, O., Casselman, A., Reppert, S.M. (2005) The two CRYs of the butterfly. *Current Biology*. **15**(23), R953–R954.

Ziegler, T.A., Forward, R.B. (2005) Larval release rhythm of the mole crab *Emerita talpoida* (Say). *The Biological Bulletin*. **209**(3), 194–203.