Università degli Studi di Ferrara

DOTTORATO DI RICERCA IN
BIOLOGIA EVOLUZIONISTICA E AMBIENTALE

CICLO XXVII

COORDINATORE Prof. GUIDO BARBUJANI

Molecular and Bioinformatic Analysis of the Circadian Clock in *Phreatichthys andruzzii*

Settore Scientifico Disciplinare BIO/05

Dottorando
Dott. NEGRINI PIETRO

Tutore
Prof. BERTOLUCCI CRISTIANO

Anni 2012/2014

TABLE OF CONTENTS

Abstract (english)

Daily cycles of light and temperature imposed by the rotation of the Earth on its axis have had a major impact on the evolution of all living organisms. Fascinating demonstrations of this fact can be seen in extreme environments such as caves or in the deep sea, where some species have evolved in complete isolation from daily light-dark cycles for millions of years, sharing a range of striking physical characters acquired by convergent evolution such as loss of eyes and pigmentation. One fundamental issue is to investigate whether these "hypogean" species still retain a functional circadian clock, which is a highly conserved self-sustaining timing system that allows organisms to anticipate daily environmental changes and is synchronized primarily by light. In this study, we have performed a comparative analysis of the circadian clock between the Somalian cavefish *Phreatichthys andruzzii*, which has evolved in perpetual darkness, and the model species *Danio rerio* (the zebrafish) that is evolved under natural daily light-dark cycles. It has been demonstrated that *P. andruzzii* retains a food-entrainable clock that is synchronized in response to regular feeding time, but does not respond to light-dark cycles. Moreover, under constant conditions, the cavefish clock oscillates with an extremely long period and also lacks normal temperature compensation. Based on these previous results, we started to analyze in detail one specific clock gene that is light-induced in zebrafish, *Period2*, where we encountered significant mutations in cavefish. We characterized the coding sequence and the genomic structure of this mutated cavefish gene and we analyzed its expression levels in comparison with zebrafish. Subsequently, for the first time in this species we performed a detailed characterization of clock and visual/non-visual photoreceptor genes as part of a complete *P. andruzzii* genome and transcriptome analysis. Our RNAseq analysis revealed a surprising phenomenon in cavefish: *P. andruzzii* mRNA sequences present an unusually high level of retained introns, leading to premature stop codons being introduced into the coding sequences of many transcripts. This mechanism may contribute to the aberrant clock phenotype of this species as well as, from a wider perspective, to other fascinating cavefish adaptations to life in constant darkness. We analyzed in detail this aberrant intron splicing phenomenon in two cavefish transcriptomes, from the brain and from a fin-derived cell line. Finally, the creation of a first *P. andruzzii* genome assembly by DNAseq analysis allowed us to map and characterize the group of visual/non-visual opsins, the main candidate genes involved in the circadian and non- photoreception

system. We used this sequence data in order to perform a comparative analysis of the levels of expression of these genes between cavefish and zebrafish. Our results suggest that the loss of photoreceptor function in cavefish may result either from mutations affecting the coding regions of the opsins, as documented in *TMT-opsin* and *Opn4m2*, or, alternatively, from mutations affecting the regulation of the expression levels of this group of photoreceptive genes.

Keywords: circadian clock, cavefish, zebrafish, *Period2* gene, transcriptome, genome, *de novo* assembly, intron splicing, visual and non-visual ospins.

Abstract (Italiano)

I cicli giornalieri di luce e temperatura imposti dalla rotazione della Terra sul suo asse hanno avuto un forte impatto sull'evoluzione di tutti gli organismi viventi. Dimostrazioni affascinanti di questo fatto si possono trovare in ambienti estremi come grotte o abissi marini, in cui alcune specie si sono evolute durante milioni di anni in completo isolamento dai cicli giornalieri di luce e buio, acquisendo un'ampia gamma di caratteri fisici come la perdita degli occhi e della pigmentazione attraverso un'evoluzione convergente. Un problema fondamentale è quello di verificare se queste specie ipogee posseggano ancora un orologio circadiano funzionale, cioè un sistema endogeno altamente conservato che permette agli organismi di anticipare i cambiamenti ambientali quotidiani e che è sincronizzato principalmente dalla luce. In questo studio, abbiamo eseguito un'analisi comparativa del sistema circadiano tra il pesce ipogeo della Somalia *Phreatichthys andruzzii*, evolutosi in buio costante, e la specie modello *Danio rerio* (lo zebrafish) la quale è normalmente esposta ai cicli giornalieri di luce e buio. È stato dimostrato che *P. andruzzii* possiede un orologio circadiano sincronizzabile in risposta a somministrazione regolare di cibo, ma non a cicli di luce-buio. Inoltre, in condizioni costanti, l'orologio di questo pesce ipogeo si esprime con un periodo estremamente lungo e manca della normale compensazione del periodo al variare della temperatura. Partendo da questi risultati precedentemente ottenuti, abbiamo analizzato in dettaglio uno specifico gene orologio la cui espressione in zebrafish è indotta dalla luce, *Period2*, in cui abbiamo scoperto mutazioni significative in *P. andruzzii*. Abbiamo caratterizzato questo gene a partire dalla sequenza codificante fino alla struttura genomica, includendo anche l'analisi dei suoi livelli di espressione confrontandoli con zebrafish. Successivamente, abbiamo sviluppato per la prima volta in questa specie una caratterizzazione dettagliata dello stato dei geni orologio e dei fotorecettori visivi e non visivi attraverso un'analisi completa del trascrittoma e del genoma di *P. andruzzii*. La RNAseq analisi ha inoltre rivelato in questa specie un fenomeno molto particolare: le sequenze di mRNA di *P. andruzzii* presentano un livello insolitamente alto di introni ritenuti al loro interno, i quali introducono codoni di stop prematuri nelle sequenze codificanti di molti trascritti. Questo fenomeno potrebbe contribuire alla spiegazione dell'orologio anormale di questa specie e, in una prospettiva più ampia, agli affascinanti adattamenti acquisiti da questo pesce ipogeo alla vita in buio costante. Abbiamo analizzato in dettaglio questi fenomeno di splicing anormale in due

trascrittomi diversi, da cervello e da linea cellulare di pinna. Infine, la creazione di un primo assemblaggio del genoma di *P. andruzzii* attraverso la DNAseq analisi ci ha permesso infine di mappare e caratterizzare il gruppo di opsine visive e non visive, i principali geni candidati coinvolti nel sistema di fotorecezione circadiana e non. Abbiamo potuto utilizzare questi dati al fine di effettuare un'analisi comparativa dei livelli di espressione di questa famiglia di geni tra *P. andruzzii* e zebrafish. I risultati ottenuti suggeriscono che la perdita della funzione dei fotorecettori in questa specie ipogea può derivare sia da mutazioni che interessano le regioni codificanti delle opsine, come documentato nei casi di *TMT-opsin* e *Opn4m2*, o, in alternativa, da mutazioni che interessano la regolazione dei livelli di espressione di questi fotorecettori.

Parole chiave: orologio circadiano, pesce ipogeo, zebrafish, gene *Period2*, trascrittoma, genoma, assemblaggio *de novo*, splicing degli introni, opsine visive e non visive.

I. Introduction

The rotation of the planet earth on its axis creates regular periods of sunlight and darkness, which have had a significant impact on the evolution of life on earth. Light is well known to be a major source of energy, as well as a fundamental signal that allows animals to interact with their environment; furthermore, most organisms regulate their physiology, behavior and biochemistry according to the time of the day.

Daily rhythms in physiology and behavior are controlled and coordinated by the circadian clock mechanism or circadian oscillator (Pittendrigh, 1993), which is an endogenous, self-sustaining time-keeping system. It is a highly conserved mechanism which continues to function even under constant conditions (so called "free-running" conditions), where it generates a rhythm with a period length of around 24 hours (Pittendrigh, 1960). For this reason it is named "circadian", from Latin *circa-diem*, that means "around a day".

One of the key features of the circadian clock is its daily resetting and synchronization by environmental signals such as light-dark cycle, food availability and temperature; these external signals are termed *zeitgebers* (from German *Zeit*: time, *geben*: givers) (Pittendrigh, 1993). Additionally, the circadian clock has been shown to be a temperature compensated mechanism (Tsuchiya et al., 2003), which means that the period length of a rhythm produced by the clock remains relatively constant over a range of temperatures.

The circadian timing system in all organisms is essentially composed of three distinct elements (Figure 1) (Menaker et al., 1978):

1. The core oscillator, that is the central cell-autonomous pacemaker which generates the circadian rhythm.
2. The input pathway, which detects the zeitgebers and entrains/resets the core oscillator.
3. The output pathway, which allows the pacemaker to control many behavioral and physiological activities.

The first reports on circadian rhythms date back to 1729, when the French scientist Jean-Jacques d'Ortous de Mairan documented the daily leaf movement of the plant *Mimosa pudica*. Later, the major properties of circadian rhythms were described by Bünning (Bünning, 1935) and Pittendrigh (Pittendrigh, 1967), using plants and insects as models. Experiments performed by Jürgen Aschhoff in humans also contributed to the emerging of the chronobiology field. The modern era of genetic analysis of the clock started in 1971 with the isolation of the *Period* (*Per*) mutant in *Drosophila melanogaster* (Konopka &

Benzer, 1971), the first clock mutant ever identified. Subsequently, genetic studies in a wide range of organisms including *Neurospora*, Cyanobacteria, *Arabidopsis*, *Drosophila* and more recently in mouse, helped to identify several clock mutants and to isolate the so-called "clock genes", the fundamental components of the core clock mechanism. As already mentioned, the basic organization of the central clock mechanism appears to be highly conserved through evolution, even though the sequences of individual clock components between certain phylogenetic groups appear to be divergent. This suggests that circadian clocks may have evolved numerous times independently during the course of evolution (van Ooijen & Millar, 2012).

In many cases, clock components serve as transcriptional regulators in the context of a transcription-translation feedback loop, which takes approximately 24 hours (circadian rhythm) to complete one cycle (Wager-Smith & Kay, 2000). Light, food and temperature changes represent the strongest entraining stimuli for the circadian clock, but how these environmental signals are detected and communicate with the core oscillator is still incompletely understood. Furthermore, the complex systemic and cell autonomous mechanisms that constitute clock output pathways are also poorly understood.

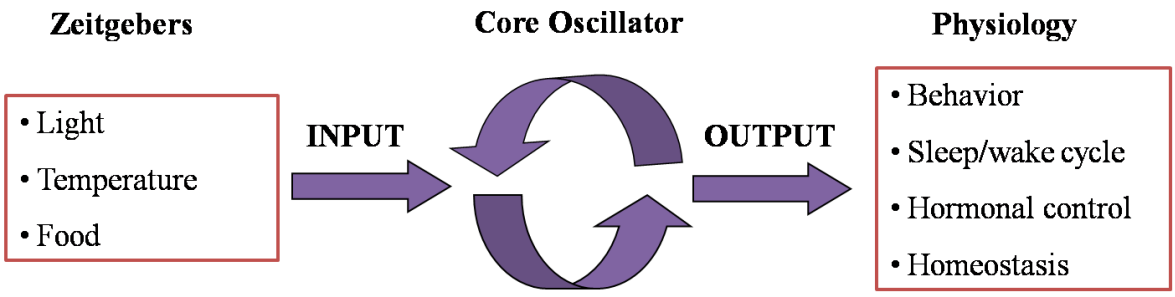

Figure 1. Schematic representation of the circadian timekeeping mechanism. By daily environmental signals, the input pathways synchronize the circadian pacemaker with natural cycles; this generates cell autonomous circadian rhythms. Through multiple output pathways, the circadian pacemaker is able to control and regulate various behavioral and physiological aspects in all living organisms.

1.1 General aspects of the circadian clock

Thanks to the early studies of the circadian clock which identified the physical localization and cellular organization of the circadian oscillator, we now know that in vertebrates the circadian clock consists of a hierarchy of multiple pacemakers. The "central" pacemakers are located in the brain and play an important role by coordinating the function of the "peripheral" clocks, which are distributed throughout tissues, organs and cells. The central pacemakers control and communicate with the peripheral clocks via complex systemic signals, which are still poorly understood. Major challenges are now to understand how all these multiple clocks communicate with each other and remain synchronized throughout the whole body.

This section will describe the general features of the circadian clock and how light and temperature can be perceived and entrain the clock.

1.1.1 Circadian pacemakers

Classically, the circadian pacemaker was considered as a single, centralized clock that regulates all the physiological rhythms in the body (Pittendrigh, 1993). For example, in *Drosophila*, pacemaker neurons located in the brain control locomotor activity rhythms. However, the spatial distribution of clock genes expression revealed the existence of multiple pacemakers located in many tissues of the fly; furthermore, studies with transgenic flies expressing luciferase or GFP under *per* or *tim* regulatory clock gene promoters, showed rhythmic expression in various tissues of this insect (Plautz et al., 1997). Rhythmic clock gene expression in all these tissues has been demonstrated even *in vitro*, underlining the existence of brain-independent autonomous clocks in many cells. For this reason, the centralized function of the circadian clock system has been updated with the notion of multiple peripheral clocks spread throughout the whole organism (Plautz et al., 1997; Yamazaki et al., 2000) and that, theoretically, every single cell or tissue can contain a circadian clock (Balsalobre et al., 1998; Tamai et al., 2003).

In mammals, the suprachiasmatic nucleus (SCN) has been shown to act as the master circadian pacemaker. Through lesioning and transplantation experiments (Moore & Eichler, 1972; Ralph et al., 1990) it has been demonstrated that the SCN is responsible for the generation and regulation of rhythms in behavior, hormonal secretion and other physiological functions (Klein & Reppert, 1991). This pacemaking ability of the SCN, as well as its response to environmental time signals, neuropeptides expression and control of

rhythms, is organized in a small network involving around 20,000 neurons (Antle & Silver, 2005) which is located in the anterior part of the hypothalamus, immediately dorsal to the optic chiasma and bilateral to the third ventricle. Within the SCN, two sub-regions have been identified, the dorsal shell and a ventral core, by their expression of clock genes (Yamaguchi et al., 2003) as well as neural peptides such as Arginine Vasopressin (AVP) and Vasoactive Intestinal Polypeptide (VIP) (Klein & Reppert, 1991). Furthermore, *in vivo* experiments revealed that rhythmic clock gene expression in the ventral core neurons is more rapidly synchronized to light-dark (LD) cycles shifts, compared to the clock gene expression from the dorsal shell neurons (Albus et al., 2005). Communication between SCN neurons involves neurotransmitters such as GABA (Liu & Reppert, 2000), VIP (Aton et al., 2005) and also gap junctions (Colwell, 2000). Another central pacemaker in vertebrates has been identified in the retina, that is responsible for controlling the local synthesis of the hormone melatonin (Tosini & Menaker, 1996), while, in non-mammalian vertebrates, a central circadian pacemaker is located in the pineal gland, driving rhythms of melatonin release. These rhythms are directly regulated by exposure to light even in cell culture (Menaker et al., 1997; Takahashi et al., 1980).

Circadian clocks do not only exist in SCN neurons, but they appear to be a property of most peripheral tissues (Dibner et al., 2010). Peripheral clock rhythms in mice and rats have been studied in many tissues such as lung, liver and skeletal muscles, using transgenic models. Explanted and cultured peripheral tissues of transgenic rats, expressing a *per1*::luciferase reporter construct, show circadian oscillations of bioluminescence that persist over several days, while explanted cultures of the SCN can continue to oscillate over a time period of at least 32 days (Yamazaki et al., 2000). Similar experiments have been performed in mice carrying a luciferase reporter gene integrated into the endogenous *Period2* locus by "knock-in" recombination (Yoo et al., 2004). In this experiments, rhythms of luciferase expression could be measured in various organs for a longer period, suggesting that peripheral clocks are more robust than supposed by the original *Period1* transgenic rat studies (Yoo et al., 2004). From these findings, the prediction would be that dysfunction of the central core clock does not necessarily inactivate the peripheral clocks, which may be instead forced to become temporally uncoupled (Nagoshi et al., 2004; M. P. Pando et al., 2002).

Peripheral clocks of mammals are not directly influenced by light, so the entrainment of peripheral clocks by light-dark cycles is indirect and occurs via the central clock in the SCN. Peripheral clocks seem to be synchronized with the central clock through a complex combination of signals, including metabolites and body temperature (Brown et al., 2002;

Damiola et al., 2000; Schibler, 2007). However, the physiological significance of peripheral clock responses independently from the SCN is not yet completely clear (Uchida et al., 2010). The situation is different in *Drosophila* and zebrafish, where the clocks in peripheral tissues and in cell cultures are directly entrainable by light (Underwood & Groos, 1982; Whitmore et al., 1998).

1.1.2 Molecular clock mechanism

In many model organisms, such as *Drosophila*, *Neurospora*, Cyanobacteria, mouse and fish, genetic analysis has identified a certain number of clock genes which are fundamental for the function of the molecular oscillator. Many of these clock genes are regulatory elements of transcription-translation feedback loops and act with a transcription repression or activation function (Wager-Smith & Kay, 2000).

In the vertebrate molecular core loop, *Clock* (*Clk*) and *Bmal* genes encode transcription factors with a basic-helix-loop-helix DNA binding domain (bHLH) (Crews & Fan, 1999) and a PAS (PER-ARNT-SIM) domain, a domain encountered in many signaling proteins where it serves as a signal sensor (Ponting & Aravind, 1997) and mediates protein–protein interactions. These two bHLH-PAS transcription factors constitute the positive elements in the vertebrate clock. Genetic experiments in mice generated by N-ethyl-N-Nitrosourea (ENU) mutagenesis, contributed to identify the *Clock* (*Clk*) gene. *Clock* mutant mice showed aberrant locomotor activity rhythms under constant darkness conditions (DD) in wheel running assays (Vitaterna et al., 1994). Subsequently, positional cloning led to the characterization of the m*Clk* locus, homologous to the *Drosophila Clock* gene (King et al., 1997). BMAL was first identified because to its capacity to interact with CLK (Gekakis et al., 1998; Hogenesch et al., 1998). In addition, gene knock-out of *Bmal* in mice caused an immediate loss of locomotor activity rhythms under constant darkness (Bunger et al., 2000). CLK and BMAL proteins heterodimerize in a transcription-activating complex; this complex binds to conserved E-box enhancer elements (5′-CACGTG-3′) that are present in the promoter region of clock-controlled genes (CCGs). These other clock genes constitute the negative regulatory elements of the core clock loop and include three *Period* genes (m*Per1*, m*Per2*, m*Per3*) and two *Cryptochrome* genes (m*Cry1*, m*Cry2*) (Dunlap, 1999).

The PER proteins are central elements of the mammalian circadian clock. They contain two PAS (PER-ARNT-SIM) domains (PAS-A and PAS-B), which can mediate homodimeric and heterodimeric mPER-mPER interactions, as well as interactions with transcription factors and kinases. These three mammalian *Period* gene homologs have been

identified by their amino acid similarity with the *Drosophila* PER protein (Shearman et al., 1997; Tei et al., 1997), and the discovery of multiple *Period* genes suggested either specialization or redundancy of function of these family members. All three PER proteins function as negative regulators of CLK and BMAL; however, in mammals, this inhibition is not as strong as in *Drosophila* (Kume et al., 1999). Systematic knock-out studies have revealed distinct roles of the *Period* genes in the circadian clock mechanism (Bae et al., 2001; Shearman et al., 2000). Whereas m*Per1* and m*Per2* seem to be essential, m*Per3* is not necessary for circadian rhythmicity (Shearman et al., 2000). Mutations in single m*Per* genes do not result in loss of circadian clock function, while double mutants of m*Per1* and m*Per2* lead to a complete loss of rhythmicity (Zheng et al., 2001). M*Per1* and m*Per2* are expressed with a circadian rhythm and rapidly induced in the SCN by administration of light pulses during the subjective night, but not during the subjective day (Shigeyoshi et al., 1997; Zylka et al., 1998). Furthermore, phosphorylation of the mPER proteins regulates the rate of turnover, ensuring a delay in their accumulation and in their negative action within the feedback loop (Maier et al., 2009). The importance of the phosphorylation sites in the PER proteins is demonstrated by a mutation affecting the phospho-acceptor site for CK1ε in the hPER2 protein, which is responsible for the familial advanced sleep syndrome (Toh et al., 2001). However, our understanding of the precise contribution of these genes to the entrainment of the clock by light and their molecular function in general, is still incomplete (Albrecht et al., 2007; Albrecht et al., 2001).

The mammalian *Cry1* and *Cry2* genes are also rhythmically expressed and can form complexes with PER proteins. Knock-out mice deficient in *Cry1* exhibit shorter free-running circadian oscillators, whereas mice deficient for *Cry2* show a longer free-running circadian oscillator (van der Horst et al., 1999; Vitaterna et al., 1999). Interestingly, double knock-out mice for m*Per1* and m*Per2* (Zheng et al., 2001), as well as double knock-out mice for m*Cry1* and m*Cry2* (van der Horst et al., 1999), show nearly the same phenotype, with an immediate arrhythmic locomotor activity after transfer from light-dark cycle (LD) to constant darkness (DD) conditions. Thus, both the cryptochrome and period proteins appear to function in combination as negative elements in the core feedback loop. Following the activation of *Per* and *Cry* gene expression by CLK:BMAL, the complexes of PERIOD (PER) and CRYPTOCHROME (CRY) proteins subsequently enter the nucleus as heterodimers and inhibit their own transcription by blocking the transactivation of the CLK:BMAL heterodimer, thereby closing the core loop (Reppert & Weaver, 2001). Thanks to the ability of the CLK:BMAL complex to directly regulate the expression of

other non-clock genes, this feedback loop can control many physiological and behavioral functions.

An additional feedback loop involves the nuclear orphan receptors REV-ERBα and RORα. REV-ERBα and ß are transcription factors belonging to the retinoic acid-related orphan receptor family. The expression of REV-ERBα is activated by CLK and BMAL and repressed by factors from the negative limb. REV-ERBα, in turn, represses the transcription of the *Bmal* gene (Preitner et al., 2002). Furthermore, RORα can compete with REV-ERBα for the binding of shared DNA-binding elements present in the promoter region of *Bmal* (Sato et al., 2004). Genes including these REV-ERB/RORE elements in their promoters are repressed by REV-ERBα, whereas RORα activates their transcription. Thus, the REV-ERBα/RORα feedback loop plays a major role by interconnecting the positive and negative limbs of the clock (Duez & Staels, 2008). The presence of this loop is predicted to confer a much higher stability on the core loop in the mammalian clock (Emery & Reppert, 2004; Preitner et al., 2002) (Figure 2). In the case of REV-ERBα, a heme group has been implicated as its ligand (Raghuram et al., 2007; Yin et al., 2007). This finding suggests that heme regulation of REV-ERBα might link the mammalian core clock with the control of metabolism (Raghuram et al., 2007). Recent results have also revealed that double knock-out mice for the REV-ERBα and ß genes show a phenotype with a strongly disrupted circadian clock function (Cho et al., 2012), and mice treated with synthetic REV-ERB ligands have a significantly altered rhythmic clock, as well as disruptions in metabolic gene expression and abnormal circadian behavior (Solt & Burris, 2012).

In addition to the core loop, a complex system of post-translational modifications of clock proteins regulates their turn over and their sub-cellular localization during the circadian cycle (Mehra et al., 2009). Furthermore, transcriptional regulation within the clock gene machinery requires the rhythmic recruitment and assembly of multi-protein complexes in a circadian time dependent manner. These events are connected with rhythmic changes in epigenetic regulation, normally carried out by chromatin modifying enzymes (Borrelli et al., 2008; Masri & Sassone-Corsi, 2010). In the case of the CLK:BMAL complex, in addition to its ability to bind to chromatin via the E-boxes in a circadian manner, CLK itself possesses an histone acetyl-transferase (HAT) activity, which is predicted to contribute to the opening of chromatin fibers and promoting rhythmic transcription of clock controlled proteins, through the histone modifications H3, K9 and K14 (Nakahata et al., 2007). In addition, the CREB binding protein (CBP) as well as p300 coactivators both have HAT function and are rhythmically recruited by the CKL:BMAL complex. However,

while p300 seems to function as a co-activator of CLK:BMAL by influencing the acetylation state of histones in a circadian fashion, the function of CBP within this process is not yet fully understood. CBP also interacts with PER2, thereby participating in the negative limb of the clock feedback loop (Etchegaray et al., 2003; Koike et al., 2012).

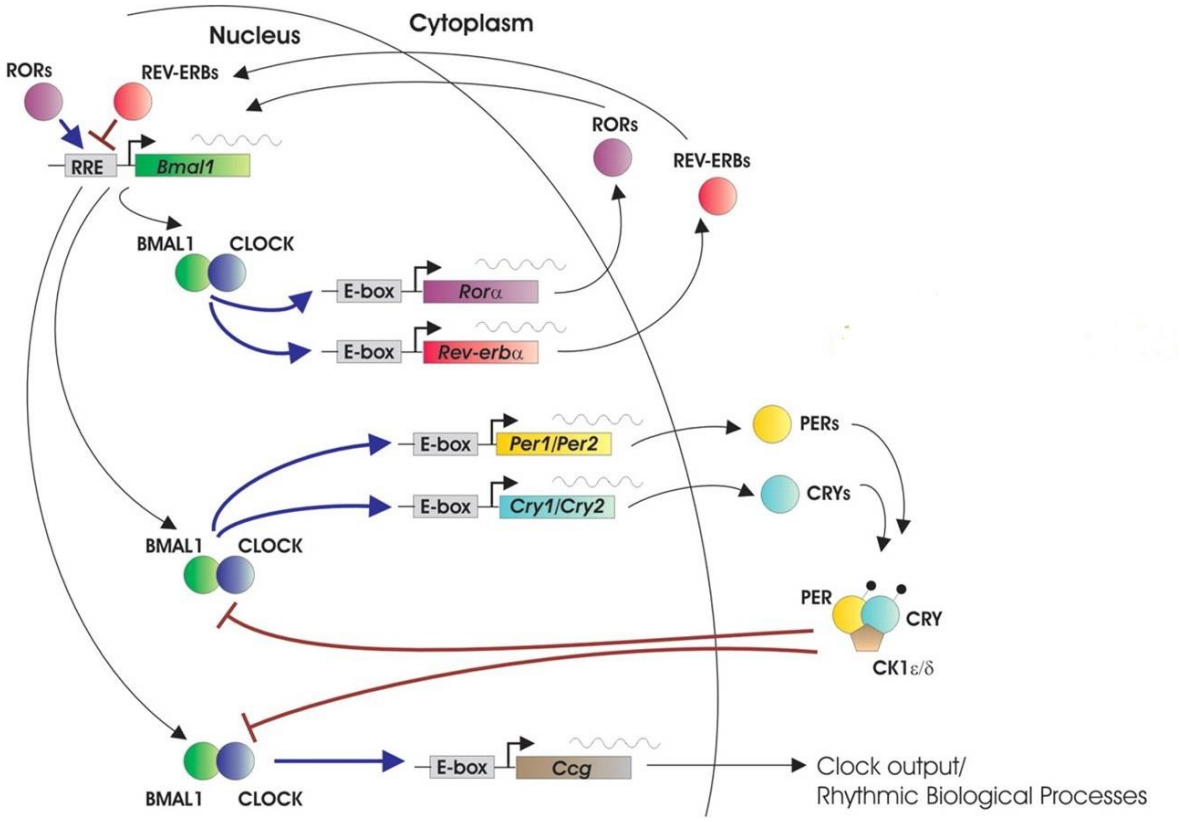

Figure 2. The core of the molecular clock mechanism in mammals. The clock is composed by two interacting loops. The core loop is formed by the positive elements CLOCK and BMAL, which heterodimerize and activate the transcription of the negative elements of the loop, the *Per* and *Cry* genes. Subsequently, PER and CRY proteins enter the nucleus as heterodimers and repress their own transcription by the inhibition of CLOCK:BMAL action. This mechanism is stabilized by a secondary loop, in which the CLOCK:BMAL complex induces the expression of *Rev-Erbα* and *Rorα*, regulating in this way the transcription of the *Bmal* gene.

Another transcriptional activator that interacts with CLK is the methyltransferase mixed lineage leukemia 1 (MLL1) protein, that specifically tri-methylates histone H3K4 in the context of circadian clock gene promoters and is also associated with activation (Katada & Sassone-Corsi, 2010). Conversely, methylation of H3K27 on circadian promoters by EZH2 is associated with CRY dependent inhibition of transcription (Etchegaray et al., 2006). Recently a histone deacetylase HDAC3 and its co-activator NCoR were also demonstrated

to play an important role in the maintenance of the epigenetic properties of circadian genes in the mouse liver (Feng et al., 2011). All these findings reveal the importance of the clock-directed epigenetic landscape, in order to generate a rhythmic clock gene expression.
The core clock mechanism of transcription-translation feedback loops is very similar in flies and mammals. There are only two main differences in the *Drosophila* clock, compared to mammals: the function of the CRY protein and the TIM protein. In *Drosophila*, in addition to its function in controlling transcriptional activity as in the case of the mammalian CRYs, the *Drosophila* CRY protein acts also as a blue-light photoreceptor that interacts with TIM to support its degradation (Hardin, 2011). In other insect species, a CRY2-like protein with a primary function as transcriptional repressor was also found. The TIM protein is essential for clock function in flies, since it forms a complex with PER and inhibits the positive elements of the feedback loop. Differently, the mammalian TIM protein is a closer homolog of the *Drosophila* gene *Timeout*, and its precise function in the circadian oscillator remains unclear (Tomioka & Matsumoto, 2010). As in mammals, phosphorylation of clock proteins also represents a key mechanism to regulate the timing of nuclear entry and protein turnover in *Drosophila*. Phosphorylation of PER by the DBT kinase affects the protein stability of PER and the nuclear entry of the PER-TIM complex (Nawathean & Rosbash, 2004).

1.1.3 Light input pathways

Regular cycles of light and darkness represent the most important zeitgebers for synchronizing the circadian clock (Roenneberg et al., 2003; Roenneberg & Foster, 1997) and for this reason the light input pathways have been extensively studied. Other zeitgebers such as feeding time (Damiola et al., 2000), temperature changes (Lahiri et al., 2005) and social interactions (Levine et al., 2002) are also very important in regulating the circadian system.
The mammalian circadian clock detects light exclusively through the retina (Foster, 1998). To entrain the circadian clock by light, rods or cones, which were for a long time considered to be the only retinal photoreceptors, are not required. Instead, the circadian clock detects light via intrinsically photosensitive retinal ganglion cells (ipRGC), a group of non-image forming photoreceptors which innervate the SCN (Berson et al., 2002) and also play important roles in the pupillary light reflex (PLR) and sleep (Berson et al., 2002; Guler et al., 2008; Hatori et al., 2008). Genetic disruption of retinal rods and cones failed to eliminate light entrainment of the circadian clock (Freedman et al., 1999; Lucas et al.,

1999), suggesting that these photoreceptive cells are essential for vision, but they do not have a direct circadian photoreceptive role. These findings were confirmed also in humans and mice which suffer blindness caused by degradation of the outer retinal layer, but are still able to synchronize their circadian clocks by exposure to LD cycles (Ebihara & Tsuji, 1980).

Light detection in the retina is mediated by opsin photoreceptors. Opsins belong to a large group of G-protein-coupled membrane receptors, normally around 350 amino acids (AA) in length. These photoreceptors consist of palisades of seven α-helical transmembrane regions enclosing a ligand binding pocket. In this pocket, the chromophore is bound via a Schiff-base linkage to a lysine residue in the seventh helix. Absorption of a photon by the chromophore leads to the photoisomerization of this molecule from a 11-cis to an all-trans state, which constitutes a key first step in opsins' response to light. Vertebrate visual pigments have a spectral sensitivity determined primarily by the structure of the opsins (Bowmaker, 2008). Beside the rod class of pigment, there are four spectrally different classes of cone pigments encoded by separate opsin gene families: a long-wavelength class (LW) sensitive from red to green spectral region (490-570nm), a middle-wavelength class (MW), a short-wavelength opsin (SW2) sensitive for blue-violet (410-490nm) and a second short-wavelength opsin (SW1) sensitive for the violet-ultraviolet region (355-440nm) (Yokoyama, 2000).

In 1998, melanopsin was discovered by Provencio and colleges (Provencio et al., 1998), who cloned this novel opsin from the dermal melanophores of *Xenopus* (Provencio et al., 2000). Knock-out of the melanopsin gene alone showed a reducing effect on entrainment of the circadian clock by light, but not a complete loss of circadian clock rhythms (Hattar et al., 2002; Ruby et al., 2002). Interestingly, a complete loss of circadian photoreception was only detected after the combined ablation of melanopsin together with the rod and cone cells (Lucas et al., 2003; Panda et al., 2003), showing that the rod and cone cells also play some supporting role in circadian photoreception (Hattar et al., 2003). The photopigment melanopsin is a typical opsin-like seven transmembrane protein bound to a vitamin A chromophore, that is in most animals a 11-cis-retinaldehyde (Sexton et al., 2012). When light activates the melanopsin, it interacts with a G-protein of the Gq family and tiggers a phototransduction cascade (Panda et al., 2005), whereby nerve impulses from the melanopsin-containing ganglion cells are transduced through their axons via the retino-hypothalamic tract (RHT) to specific brain targets, including the olivary pretectal nucleus (OPN), that is a centre responsible for pupil size, and the SCN (Hattar et al., 2002). The ipRGCs influence these target regions by releasing from their pre-synapses the

neurotransmitters glutamate and pituitary adenylate cyclase activating polypeptide (PACAP) (Hannibal et al., 2004). The release of these neurotransmitters activates various signalling pathways, initiates chromatin re-modelling (Crosio et al., 2000) as well as inducing clock genes (Zylka et al., 1998) and immediate-early genes (Dziema et al., 2003). However, the best-studied light activated pathway in target SCN neurons is the extracellular signal-regulated kinase (ERK) pathway. Activated ERK phosphorylates cyclic adenosine monophosphate (cAMP) response element binding protein (CREB) (Dziema et al., 2003). In turn, phosphorylated CREB binds to cAMP response elements (CRE) located in the promoters of target genes and activates their transcription.

1.1.4 Temperature input pathway

In most organisms, environmental temperature cycles can also act as an efficient zeitgeber for the circadian clock. However in mammals, which are homeotherms being able to regulate their body temperature independently from environmental changes, external environmental temperature cycles represent only a very weak entraining stimuli (Refinetti, 2010). Nevertheless, core body temperature in mammals is rhythmic and under circadian control by the SCN master clock. Peripheral oscillators in various organs such as liver, kidney and lung are highly sensitive to temperature cycles (Brown et al., 2002; Buhr et al., 2010). Temperature pulses can effectively reset these oscillators and temperature cycles that match circadian body temperature rhythms can efficiently entrain peripheral clocks.

Temperature appears to regulate peripheral clocks via the transcription factor heat-shock factor 1 (HSF1). It is documented that expression of HSF1 oscillates with a circadian rhythm in the liver of mice, which can be entrained by temperature cycles. Furthermore, the blocking of HSF1 protein by a specific chemical inhibitor also blocks temperature-induced resetting of liver and lung cell culture clocks, implicating a key role for this protein in temperature entrainment of mammalian clocks (Buhr et al., 2010).

Recently, a subset of central pacemaker neurons have been shown to contribute to temperature-induced synchronization of locomotor activity in *Drosophila* (Glaser & Stanewsky, 2005, 2007; Sehadova et al., 2009). These cells depend on input from peripheral tissues (Picot et al., 2009; Sehadova et al., 2009). Members of the transient receptor potential (TRP) channel family include a well-established class of temperature detectors (Ramsey et al., 2006), and consistently a role for *Drosophila trpA1* in temperature synchronization of circadian activity was recently identified (Y. Lee & Montell, 2013).

1.1.5 Output pathways

Most animals show circadian rhythms in many, diverse physiological processes such as metabolic regulation, hormonal releases (Dickmeis, 2009) and muscle tone, as well as in behaviour such as locomotor activity, alertness, and feeding (Dunlap & Loros, 2004). It is assumed that circadian regulation of these physiological and behavioral processes can confer survival advantages. Therefore, other fundamental components of the circadian timing system are the output pathways, through which various physiological and behavioral aspects are linked with the core clock machinery. Clock outputs allow the whole organism to adapt to environmental conditions and act also at the cellular level, where peripheral clocks are able to differently regulate the expression of non-clock genes (Schibler & Sassone-Corsi, 2002). For example, in mammals, circadian rhythmicity of sleep-wake cycles and hormone production is linked to the functionality of the SCN (Reppert & Weaver, 2001). The SCN controls the nocturnal synthesis of melatonin in the mammalian pineal gland via indirect adrenergic innervations and in turn, circadian rhythms of circulating melatonin also affect many aspects of physiology (Pevet & Challet, 2011).
Another important clock regulatory target is the cell cycle. Circadian-dependent cell cycle regulation has been clearly documented in many different peripheral tissues such as skin, intestine, liver, gut and the heart (Bjarnason & Jordan, 2000), where there is evidence that the circadian clock can regulate steps in DNA synthesis and mitosis. Daily timing of cell cycle processes has been predicted to represent a strategy which can avoid the potentially damaging effects of UV exposure to sunlight (Mori et al., 1996).

1.2 Zebrafish as a circadian clock model system

The zebrafish (*Danio rerio*) (Figure 3) is a subtropical fish belonging to the *Cyprinidae* family and is considered one of the most important models for studying vertebrate embryogenesis, early development and toxicology. In the last decade zebrafish has also become an important genetic model for studying the vertebrate circadian timing system. Zebrafish are easy to maintain and breed, they show high fecundity, short generation time and their embryonic development occurs outside the mother. The transparency of the embryos makes possible the observation of individual cells during the earliest stages of development as well as embryogenesis; for this reason zebrafish represent an ideal model

for *in vivo* imaging of various biological processes (Dahm & Geisler, 2006). Large scale forward genetic screens have been successfully applied, the zebrafish genome is fully sequenced and many tools are available for generating transgenic lines (Chatterjee & Lufkin, 2012; Haffter et al., 1996; Mullins & Nusslein-Volhard, 1993). Zebrafish also show some a fascinating capacity to regenerate many tissues upon injury, including the fin, skin, heart and brain. This aspect has become an important focus of biomedical research and makes this model very attractive to study the origin and potential therapies for cardiac and neurodegenerative diseases, as well as cancer (Brittijn et al., 2009; Ingham, 2009).

Figure 3. The cyprinid *Danio rerio*. The zebrafish (*D. rerio*) is a tropical freshwater fish belonging to the Cyprinidae family. It is considered an important model organism in many research fields, including the study of the circadian timing system in vertebrates.

Zebrafish in the last years has also become an attractive species for the study of the origin and function of the vertebrate circadian clock. Grieg Cahill originally studied melatonin production in cultured zebrafish pineal and retina, detecting circadian rhythms of melatonin production in pineal cultures for up to five days under constant conditions (Cahill, 1996). At a stage when the molecular components of the vertebrate clock were completely unknown, it was predicted that, combined with large scale forward genetic analysis, this species would represent an attractive model for identifying vertebrate circadian clock genes as well as genes related to clock input/output pathways (Cahill, 2002; Haffter et al., 1996). However, early studies of clock gene expression in zebrafish revealed an additional attractive feature of this model for studying the circadian timing system in vertebrates. In particular, it was shown that zebrafish cells and tissues are intrinsically light responsive, meaning that direct light exposure can reset and synchronize their peripheral circadian clocks. Light sensitive oscillators in peripheral tissues imply that the zebrafish circadian system is organized as a distributed set of pacemakers which are independently entrained by light (M. P. Pando et al., 2001; Whitmore et al., 1998).

Furthermore, this observation predicted the widespread expression of photoreceptors in zebrafish tissues and cell lines, thus making zebrafish an important model to study the regulation of the vertebrate circadian clock by light.

1.2.1 Zebrafish circadian clock genes

After the characterization of the clock genes in mouse by cloning (King et al., 1997), rapid progresses were made in the exploration of the core clock mechanisms in vertebrates, including zebrafish (Shearman et al., 1999). During the evolution of the teleost group, a genome duplication event occurred and led to many cases of gene duplications, including several genes involved in circadian clock regulation (Meyer & Schartl, 1999; Postlethwait et al., 1998; Wang, 2008).

Many zebrafish clock genes have been isolated either by sequence homology with their mammalian counterparts or by two-hybrid screens for interacting partners of the CLOCK protein. In detail, the following zebrafish clock genes have been so far identified: three *Clock* genes (*Clk1a*, *Clk1b*, *Clk2*) (Whitmore et al., 1998), three *Bmal* genes (*Bmal1a*, *Bmal1b*, *Bmal2*) (Cermakian et al., 2000), four *Period* genes (*Per1a*, *Per1b*, *Per2*, *Per3*) (Delaunay et al., 2003; Vallone et al., 2004), six *Cryptochrome* genes (*Cry1a*, *Cry1b*, *Cry2a*, *Cry2b*, *Cry3*, *Cry4*) (Kobayashi et al., 2000), three *Ror* genes (*Rorα*, *Rorß*, *Rorγ*) (Flores et al., 2007) and one *Rev-erbα* gene (Kakizawa et al., 2007). *Cryptochrome1a*, *1b*, *2a* and *2b* share high sequence homology with the mammalian m*Cry1* gene and they also repress CLK:BMAL complex directed transcriptional activation, while *Cry3* and *Cry4* do not. *Cry3* shares sequence homology with the mammalian *Crys*, whereas *Cry4* shows higher homology with *Drosophila Cry* (Kobayashi et al., 2000). Interestingly, the zebrafish *Clk* gene exhibits a rhythmic expression profile while the mammalian temporal expression of *Clk* is relatively constant (Whitmore et al., 1998). Furthermore, expression of a broad range of genes, including the DNA-repair enzyme gene *Cry5* (alias *6-4 photolyase*), the key clock genes *Cry1a*, *Per2* and numerous other genes with different cellular functions, are directly activated by light exposure of zebrafish tissues as well as in zebrafish cell lines (Gavriouchkina et al., 2010; Weger et al., 2011).

1.2.2 Zebrafish light input pathway

In non-mammalian vertebrates including fish, specialized photoreceptive organs that develop from the embryonic forebrain, such as the eyes, the pineal gland and deep brain

photoreceptors, play an important role in circadian photoreception (Peirson et al., 2009) (Figure 4). However, most zebrafish tissues and cells are directly light responsive (Kaneko et al., 2006; Whitmore et al., 2000), exhibiting then a general tissue photosensitivity and thereby making this species a powerful model to study light input to the vertebrate clock (Vatine et al., 2009) and peripheral clock regulation in general (Vallone et al., 2004).

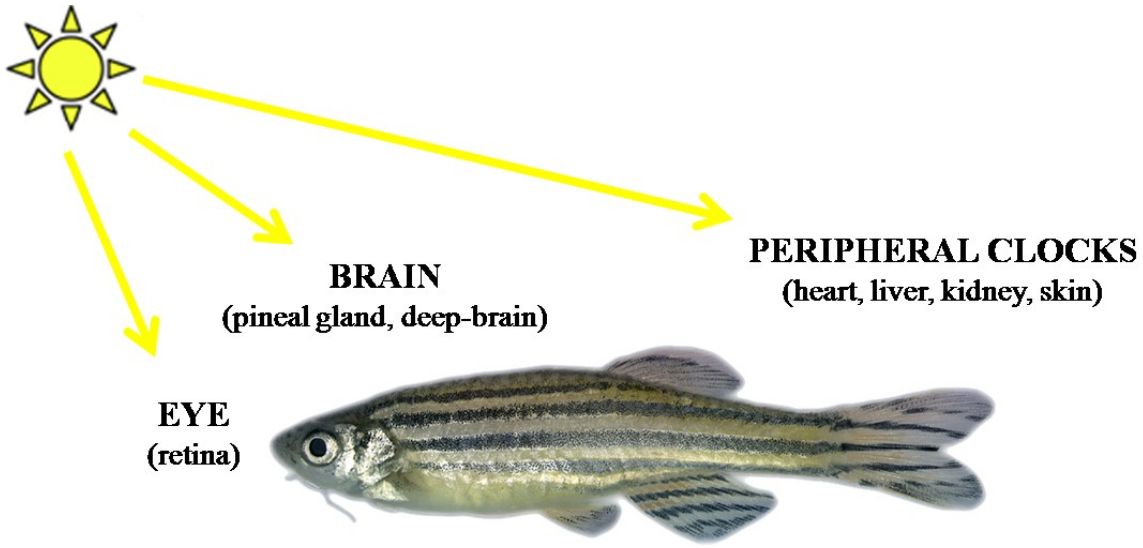

Figure 4. Main photoreceptive organs in non-mammalian vertebrates. In non-mammalian species the pineal complex, as well as deep brain structures, contain photoreceptors. Direct photoreception in isolated tissues, besides the classical photoreceptors in the retina of the eye, have been also described. Moreover, in zebrafish, the circadian clocks in peripheral tissues have been shown to be directly light entrainable.

A fundamental discovery in zebrafish circadian clock studies was that the exposure of zebrafish explanted tissues or cell lines to LD cycles, triggers robust rhythmic expression of clock genes. Subsequently, under constant conditions, rhythmicity is lost over the course of several days, due to progressive asynchrony of the individual cell clocks (Carr & Whitmore, 2005; Kaneko et al., 2006; M. P. Pando & Sassone- Corsi, 2002; P. Pando, Cermakian, Sassone-Corsi, 2001; Whitmore et al., 2000). The precise nature of the widely expressed photoreceptors still remains unclear, however three possible candidates have been implicated:

1. The zebrafish *Cryptochromes* (*Crys*) act mainly as repressors of the CLK:BMAL complex and share high homology with their mammalian counterparts. However, one zebrafish *Cry*, namely *Cry4*, shows higher homology to the *Drosophila Cry* (Kobayashi et

al., 2000), which functions as a blue light photoreceptor. For this reason, it has been speculated that zebrafish *Cry4* might also function as a photoreceptor (Kobayashi et al., 2000; Ozturk et al., 2009).

2. H_2O_2 production by a phototransducing flavin-containing oxidase is considered to be a potential light detecting mechanism, given the fact that light exposure triggers increases in intracellular levels of H_2O_2 in different model systems (Hirayama et al., 2007; Hockberger et al., 1999).

3. Opsins, as well as extra retinal non-visual opsins, are also candidates for peripheral tissue photoreceptors. While the visual opsins are expressed exclusively in the retina and contribute to image formation, recent studies have revealed an impressive diversity of opsin genes in fish species, many of which are also expressed in peripheral tissues, such as melanopsin (*Opn4m2*), teleost multiple opsin (*TMT-opsin*) and exorhodopsin (Bellingham et al., 2002; Davies et al., 2011; Mano et al., 1999; Moutsaki et al., 2003; Peirson et al., 2009; Pierce et al., 2008). *TMT-opsin*, which functions as a non-visual opsin, is expressed in many neural and non-neural zebrafish tissues, including zebrafish embryonic cell lines. Thus, *TMT-opsin* has been proposed as one candidate for the circadian clock photoreceptor in zebrafish peripheral clocks (Moutsaki et al., 2003). Interestingly, it was shown that the expression level of the long-wavelength cone opsin *Opn1lw1*, which belongs to the visual opsins, shows a day-night rhythm and this pattern of fluctuation persists also in zebrafish that are kept in constant darkness, suggesting an involvement in circadian clock function (P. Li et al., 2008).

Some light responsive enhancer elements have been identified within the promoter regions of light inducible clock genes, even if, at present, the precise mechanism of the signal transduction cascade and the transcriptional control of the light input pathways remain unknown. In the promoter region of the light induced clock gene *Per2*, E- and D-box enhancer elements have been identified, which are shown to be necessary and sufficient for light-regulated gene expression (Vatine et al., 2009). Recently, a detailed promoter study of the light-regulated gene *Cry1a* also showed that D-box enhancer elements serve as primary light responsive elements (Mracek et al., 2012). D-box binding factors were first identified in mammalian cells, including the transcription factor E4BP4, that serves as a repressor, and the PAR domain factors (TEF, DBP and HLF), which function as activators (Cowell, 2002; Gachon, 2007). In zebrafish, 12 different factors sharing high homology with the mammalian D-box binding factors have been cloned and characterized. These factors are predicted to play a key role as nuclear targets of the light driven signaling pathway (Ben Moshe et al., 2010).

1.2.3 Zebrafish temperature input pathway

Zebrafish can be used as a model also for studying the effects of body temperature changes on the circadian clock. It has been demonstrated that small shifts of temperature (±2°C) can entrain rhythmic expression of zebrafish circadian clock genes such as *Per1b*, *Cry2a* and *Cry3* (Lahiri et al., 2005; Lopez-Olmeda & Sanchez-Vazquez, 2009). It was also shown that the circadian clock is temperature compensated, meaning that the period length remains constant over a range of temperatures. Moreover, temperature changes strongly influence the amplitude of rhythmic circadian clock-driven gene expression during and following entrainment by LD-cycles, and temperature-dependent changes in the expression levels, phosphorylation and function of the protein CLK was also demonstrated (Kaneko & Cahill, 2005; Lahiri et al., 2005). Given the fact that zebrafish peripheral clocks are directly entrainable both by light and temperature, this species represents an ideal genetic clock model to study combined effects of temperature and light on the clock.

1.2.4 The role of the SCN and pineal gland in zebrafish

The functional relevance of the SCN in zebrafish, regarding circadian clock regulation, is still unclear (Kaneko et al., 2006). Given that most zebrafish tissues, cells and organs have directly light-entrainable endogenous oscillators, it has been speculated that central pacemaker function may not even be required (Kaneko et al., 2006; Whitmore et al., 2000; Whitmore et al., 1998). The pineal gland has been proposed as one structure that could function as a central pacemaker in zebrafish, given its role as a key endocrine system element, directing the rhythmic synthesis of melatonin (Falcon, 1999). As for most non-mammalian vertebrates, the zebrafish pineal organ also contains an intrinsic circadian oscillator that is directly regulated by light (Falcon et al., 2007; Korf et al., 1998). Additionally, circadian rhythms in zebrafish continue even in the presence of a genetically disrupted SCN, providing even more evidence that the pineal gland could act a central independent pacemaker in zebrafish (Noche, 2011).

While in non-mammalian vertebrates the pineal gland includes all three functional elements of the clock (input, pacemaker and output), in mammals this system has become more specialized. The melatonin production in the mammalian pineal gland is indirectly controlled by the SCN, and photoreception appears only via retinal photoreceptors (Falcon, 1999). Melatonin, also known as N-acetyl-5-methoxytryptamine, is synthesized in four enzymatic steps from the essential dietary amino acid tryptophan. Hydroxylation of

tryptophan is catalyzed by tryptophan hydroxylase (TPH), and the following decarboxylation by aromatic acid decarboxylase (AAD) results in the production of serotonin. The subsequent two steps, N-acetylation by AANAT2 (arylalkylamine-N-acetyltransferase) and O-methylation by HIOMT (hydroxyindole-O-methyltransferase), occur mainly in the pineal gland and in retinal photoreceptors (Ganguly et al., 2001). AANAT2 particularly important since it directly controls the rate of melatonin production, meaning that the higher melatonin levels at night are directly coupled with increased AANAT2 activity (Ziv et al., 2007).

The controlling mechanism of gene expression in the pineal gland has been extensively studied due to the importance of this structure. Promoter studies using a transient transgenic promoter assay or transfection of cell lines have revealed a cluster of enhancer elements lying downstream of the *aanat2* gene, which are critical in confering its specific temporal and spatial pattern of expression (Appelbaum et al., 2004). This cluster, known as pineal-restrictive-downstream-module (PRDM), is composed of three photoreceptor conserved elements (PCEs) and an E-box, the regulatory target of the circadian clock transcription factors CLK and BMAL. The PCEs also represent a target for the homeobox gene, *otx5* (orthodenticlerelated homeobox 5) which determines pineal specific gene expression (Appelbaum et al., 2006).

1.2.5 Circadian clock outputs in zebrafish

Many aspects of the physiology and behavior of zebrafish are regulated by the circadian clock, as revealed by several physiological, behavioral and molecular studies. As already mentioned, the zebrafish retina and pineal gland contain photosensitive circadian oscillators which are able to drive robust rhythms of melatonin synthesis. Subsequently, these melatonin rhythms are responsible for the regulation of diverse physiological functions (Cahill, 1996). It is also documented that cell cycle rhythm in zebrafish is clock regulated and direct exposure to visible light appears to play a key role in timing of cell cycle progression (Dekens et al., 2003). Diurnal locomotor activity rhythms under LD cycles have also been described in adult zebrafish, thus appearing to represent another robust circadian clock output (Hurd et al., 1998). These diurnal locomotor rhythms of activity are interestingly suppressed by constant light (LL) conditions (Lopez-Olmeda et al., 2010).

1.3 Cavefish as a new circadian clock model system

Many fish species have remained completely isolated from the day-night cycle for millions of years, in extreme environments such as subterranean caves (Colli et al., 2009); these species show a convergent evolution with a range of physical "troglomorphic" properties, such as eye degeneration during early development, complete depigmentation or reduced metabolism. They potentially represent powerful models for studying the evolution of the circadian clock under extreme environmental conditions. Many aspects of cavefish biology still remain incompletely understood; for example, how does evolution in the dark affect the regulation of the circadian clock? Do these animals even retain a "normal" circadian clock?

1.3.1 Life in the darkness: special features of cave inhabitants

Caves are normally based on soluble rock formations including limestone, dolomite or gypsum (Gunn, 2004) and typically contain interconnected water bodies such as springs or streams. Usually water systems in caves remain connected to the surface; in this way water exchange still takes place. At their deepest levels, caves can be completely isolated from environmental signals such as light and, in some cases, have remained undisturbed for millions of years, representing hostile environments with low nutrient availability because of the lack of primary production by photosynthesis (D. C. Culver, 1982). However, water can connect certain cave systems with the outside world, carrying food for example by seasonal flooding events or streams (Jeffery, 2001; Timmermann & Plath, 2009). Another source of nutrition in some cave systems is represented by roosting bats that fly every day out from the cave and then return to roost after feeding.

Cave inhabitants present characteristic changes in morphology, mainly due to the absence of light, which are termed as *troglomorphisms* and have been separated into constructive and regressive traits (Hecht et al., 1988; Pipan & Culver, 2012). Constructive traits are normally correlated with the enhanced development of non-visual sense organs, such as taste buds (Varatharasan et al., 2009), lateral line systems (Yoshizawa & Jeffery, 2011), changes in metabolic rate (Wilhelm et al., 2006) as well as in body size and posture (P. D. Culver et al., 2000; Trontelj et al., 2012). In contrast to constructive traits, regressive traits typically represent loss of an organ or function. The most obvious and severe regressive trait in cave inhabitants is the loss of eyes and pigmentation (M. E. Protas et al., 2006;

Wilkens, 2010; Wilkens & Strecker, 2003), even though, unexpectedly, a large number of cave species do not show these troglomorphic traits. Recently, it has been hypothesized that the lack of troglomorphic traits may be characteristic of organisms that have only recently occupied the cave environments, while the troglomorphic species instead reflect the long period of evolution required to acquire such traits (Pipan & Culver, 2012; Trontelj et al., 2012).

As already mentioned, one of the most prominent regressive trait in cave animals is the loss of eyes, a phenomenon intensively studied during the last decade. Progressive accumulation of mutations in eye-determining genes, after loss of the evolutionary pressure to retain visual perception, is widely accepted as the main cause of this phenotype (Wilkens, 2011). However, some believe that eyes loss might also occur by positive selection of mutations that, in conjunction with eyes loss, could also enhance the survival of the organism in the cave environment through pleiotropic effects on other sensory systems (Yamamoto et al., 2009).

In many cases, eyeless cave species have surface-living ancestors of the same or closely related species that show normal eye development, allowing comparative studies between the cave phenotype, considered as a natural mutant, and its normal surface-dwelling counterpart. Remarkably, in most cave animals eyes do form during early development, but progressively degenerate during later development, never gaining functionality. However, the developmental steps that lead to the eye loss have not yet been completely characterized at the molecular level. Development and subsequent regression of the eyes has been observed and studied in various animals, for example in the salamander *Proteus anguinus* (Hawes, 1945), the planthopper *Oliarus polyphemus* (Wessel et al., 2013), the amphipod *Gammarus troglophilus* (Aspiras et al., 2012), the crustacean *Monolistra monstruosa* (M. Protas & Jeffery, 2012), the catfish *Rhamdia laticauda* (Wilkens, 2001), the barb *Garra barraimiae* (Kruckenhauser et al., 2011; Timmermann & Plath, 2009), the cave tetra *Astyanax mexicanus* (Jeffery, 2001), and the Somalian cavefish *Phreatichthys andruzzii* (Ercolini & Berti, 1975; Stemmer at al., 2015).

1.3.2 *Astyanax mexicanus*: a teleost cavefish model

Around 86 different cave dwelling teleost species have been described worldwide (Proudlove, 2006), a heterogeneous group that share a set of convergent “regressive” phenotypes. It has been speculated that these characters might represent key adaptations to life in these extreme environments, or might either reflect the loss of selective pressure

inside the cave environments. Thus cavefish species provide a great opportunity to study convergent and parallel evolution of regressive traits, developed under environmental conditions that contrast with those of their epigean (surface-dwelling) ancestors (Jeffery, 2009). The only limiting factor to study cave animals is the difficulty to culture many of these species in the laboratory (Yamamoto, 2004).

Up to now, the best-studied cavefish is the Mexican cave tetra *Astyanax mexicanus*. It is a fresh-water fish belonging to the *Characidae* family, order *Characiformes*. Many ipogean cavefish populations (Figure 5B) and one surface epigean population of *Astyanax* have been documented (Figure 5A) (Dowling et al., 2002; Jeffery, 2009). While the surface population mainly inhabit northeastern Mexico and southern Texas, the cave populations are mostly found in the Sierra de El Abra limestone formation in Mexico (Hecht et al., 1988). The troglomorphic phenotypes of *Astyanax* cavefish populations are highly influenced by how long they have been isolated in their cave environment, and so are highly variable (Strecker et al., 2012). For example, while Piedras, Pachón and Curva cavefish have completely lost their eyes and body pigmentation, in other populations like Molino and La Cuerva intermediate or mixed variants of these traits can be found (Jeffery, 2009; Wilkens & Strecker, 2003).

Astyanax is a very well-studied cavefish model, mainly because of its suitability as a laboratory animal. This species has a short generation time (approximately 4 to 6 months), can be raised on a simple diet and spawns frequently under standard laboratory conditions (Jeffery, 2009; Wilkens, 2010). The cavefish forms of *Astyanax* and the surface form are fully inter-fertile (Sadoglu, 1957), meaning that they can be crossed with each other for genetic complementation analysis (Borowsky, 2008; Wilkens, 1971; Wilkens & Strecker, 2003), and the cavefish forms can be compared with their surface fish con-specifics in the same way as mutants are compared with the wild-type phenotypes (Jeffery, 2009). Furthermore, many molecular techniques available for other teleosts species can also be applied to *Astyanax*, such as molecular cloning, mRNA injection and *in situ* hybridisation with embryos, Quantitative Trait Loci (QTL) mapping as well as transcriptome and genome sequencing analysis.

Interestingly, the loss of pigmentation evolved in three different populations of *Astyanax* (Pachón, Molino and Japonés) via the same genetic pathway, where a loss-of-function mutation of the gene *oca2* was consistently observed (M. E. Protas et al., 2006). This cavefish model has proved to be also particularly useful for studying the evolution of eye degeneration.

Figure 5. Epigean (surface) and ipogean (cave) phenotypes of *Astyanax mexicanus*. (A): surface form of *A. mexicanus* with full-sized eye and pigmentation. (B): cave form of *A. mexicanus* (Pachón) showing loss of eyes and pigmentation. (Sources: modified from McGaugh et al., 2014).

Genetic analysis of *Astyanax* QTLs indicated that the loss of eyes is caused by several mutations, and that these mutations differ between cave populations (Gross et al., 2008). Interestingly, crosses between different cavefish populations result in offspring with intermediate eye size in the F1 generation, and give a distribution of eye size in the F2 generation, from complete absence to a normal wild-type eye, strengthening the hypothesis that cavefish located in different caves have evolved independently from each other and do not automatically carry the same mutation (Borowsky, 2008).

By transplantation experiments of lenses from surface fish to cavefish eye, and vice versa, it was suggested that the lens in the eye of *Astyanax* plays a central role in eye degradation (Yamamoto & Jeffery, 2000). These studies revealed that the lens and retina undergo massive apoptosis, suggesting an active loss of multiple regulatory pathways. The transplantation experiments demonstrated that a surface lens placode transplanted into a cavefish of the same developmental stage, leads to a normal-sized eye. Conversely, a lens placode of a cavefish transplanted into a surface fish leads to a cave-like eye phenotype (Romero et al., 2003). It has been suggested that a key difference between cave and surface fish eye development is due to increased *Sonic Hedgehog* (*Shh*) signalling at the embryonic midline (Yamamoto et al., 2004).

In conclusion, *Astyanax mexicanus*, as well as cavefish species in general, are fascinating models to study the consequences of complete isolation from the day-night cycle on many

anatomical traits, such as eye development and loss of melanin pigment. The work that will be subsequently presented in this thesis, attempts to address the evolution of the circadian timing system and non-visual photoreception mechanisms in an extreme cave environment. Another cavefish species has been used as a model for our studies, namely the Somalian cavefish *Phreatichthys andruzzii*. The next section will give a general overview of this species and its importance for our research aims.

1.3.3 The Somalian cavefish *Phreatichthys andruzzii*

The Somalian cavefish *Phreatichthys andruzzii* (Ercolini & Berti, 1975) (Figure 6) is a tropical cyprinid and an inhabitant of the subterranean water systems in the Bud-Bud region of Somalia, under the central Somalian dessert, whose ancestors probably adapt to a subterranean life in the phreatic waters due to climate changes. Its habitat is formed by large phreatic layers that develop in Eocene horizontal limestone rock formations, which are accessible only through sink holes and shallow wells.

Figure 6. The Somalian cavefish *Phreatichthys andruzzii*. This extreme animal model shows complete eye loss, no pigmentation, no scales, as well as strongly reduced metabolism and oxygen consumption (Sources: modified from Cavallari et al., 2011).

The main reason why we decided to study *Phreatichthys andruzzii*, and not the better-known *Astyanax mexicanus*, is that the Somalian cavefish presents a more extreme troglomorphic phenotype than the mexican cavefish *Astyanax*. The eye degeneration is much more rapid (Berti et al., 2001), while many *Astyanax* populations do not always show complete eye degeneration and, in some cases, they still have rudimental eyes. In *P. andruzzii,* normal formation of the optic cup has been reported around 22 hours post

fertilization and growth of the retinal layers has been shown between 32 and 36 hours post fertilization. Around the same time, “autolytic vacuoles” appear and increase in number, until 46 hours post fertilization, while the size of the lens decreases. This is followed by retinal degeneration and by 72 hours post fertilization, only a small cyst of cells is left that subsequently sinks into the surrounding tissue. Finally, the adult *P. andruzzii* presents complete anophthalmy and a reduction in size of associated brain structures (Berti et al., 2001). This is probably reflects the fact that the Somalian cavefish has evolved in its completely dark and constant environment for around 1.4-2.6 million years, approximately 1 million years longer than *Astyanax mexicanus*.

Interestingly, *P. andruzzii* still shows a notable photic sensitivity in behavioral tests, where it tends to avoid illuminated areas in favor to complete dark areas. A recent study indicated that two photopigments, rhodopsin (521 nm) and exorhodopsin (520 nm), could be involved in the observed behavioral spectral sensitivity (Tarttelin et al., 2012). Another reason of interest in this species, is that *P. andruzzii* belongs to the *Cyprinidae* family, and represents a close relative to zebrafish. This enables the use of many tools, experimental approaches and information gained from zebrafish studies. The only limitation of this species, is that the fertilization and breeding rate is not as efficient as for *Astyanax* in the laboratory environment; reproduction can only be induced by injection of gonadotropic hormones and even then, it is not as efficient as in other species such as the zebrafish. For this reason, cell cultures of cavefish cells derived from clipped-fins of *P. andruzzii* are often used as experimental tools.

Our group has previously performed a detailed molecular characterization of the *P. andruzzii* circadian clock (Cavallari et al., 2011). Clock genes have been identified and partially or completely cloned by sequence homology with zebrafish. The clock genes identified so far are as follows: seven *Cryptochromes* (*Cry1a*, *Cry1b*, *Cry2a*, *Cry2b*, *Cry3*, *Cry4, Cry5*), four *Period* genes (*Per1a*, *Per1b*, *Per2*, *Per3*), three *Clock* genes (*Clk1*, *Clk2*, *Clk3*), three *Bmal* genes (*Bmal1a*, *Bmal1b*, *Bmal2*), one *Rev-erbα* gene and two *Rorα* genes (paralog a, b). To date, the most important finding in *P. andruzzii* studies is that this species possesses a clock mechanism that does not respond to light, but rather can be entrained by regular pulses of food. Furthermore, in cavefish cell lines derived from adult fin clips, an abnormal clock mechanism has been observed, which oscillates with an extremely long period of more than 40 hours (so-called “infradian” period) and also lacks normal temperature compensation (Cavallari et al., 2011). Several mutated genes have been documented and predicted to be linked to this abnormal clock phenotype, such as the

extra-retinal circadian photoreceptors melanopsin *Opn4m2* and *TMT-opsin* (Cavallari et al., 2011).

1.4 Aim of the study

The fundamental task that guides our studies on cavefish is to try to understand what are the consequence for the circadian timing system of evolution during isolation from the day-night cycle over millions of years. For this purpose, we selected the blind cavefish *Phreatichthys andruzzii* as our model, in comparison with the classical surface-dwelling and close-relative model species *Danio rerio* (the zebrafish), in order to investigate the molecular, cellular and anatomical adaptations to life in constant darkness. Previous findings from our research group showed that *P. andruzzii* possesses an abnormal clock mechanism which oscillates with an infradian rhythm longer than 40 hours; moreover, its clock is affected by a lack of a normal light input pathways as well as of normal temperature compensation (Cavallari et al., 2011). These results represent the background for the study presented here, which is characterized by three major goals that we have tried to tackle in order to elucidate our understanding of the complex circadian timing system of *P. andruzzii*:

1. Our first aim was the detailed analysis of the *Period2* gene, a key clock gene that is light-induced in zebrafish and shows significant mutations in *P. andruzzii*, including the insertion of a transposable element in the coding sequence and a premature stop codon in the C-terminal region of the sequence. We initially developed a complete map of this gene, starting from the coding sequence up to the genomic structure. Subsequently, we tested the presence of the transposable element *in vivo* and measured *Per2/Cry1a* mRNA expression levels in a comparative analysis with zebrafish.
2. By powerful Next Generation Sequencing (NGS) approaches, we performed the first transcriptomic and genomic sequence analysis of *P. andruzzii* with the initial goal of characterizing in more detail all clock and photoreceptor genes in this species. Specifically, we obtained two different cavefish transcriptome sequences, from the brain and from the fin-derived CF-1 cell line, and a first draft whole genome sequence. Our bioinformatic analysis aimed to provide a detailed dataset of full-length sequences and to identify

particular mutations which could explain in more detail the mechanisms that underlie the aberrant clock phenotype of this species.

3. Through the assembly of our first *P. andruzzii* genome draft sequence, we mapped and characterized the group of visual/non-visual opsins, the main candidate genes involved in the circadian and non- visual photoreception system. We utilized the sequences to perform a comparative analysis of the expression levels of these genes between cavefish and zebrafish. Thus, our goal was to improve our understanding of the photoreceptor function loss in *P. andruzzii* and, from a broader perspective, to study the molecular pathways in response to light.

II. Materials and Methods

2.1 Animals

Somalian cavefish (*P. andruzzii*) were collected from different sites around the locality of Bud-Bud (04°11´19´´N-46°28´27´E) during the years 1968-1982.
The fish were transferred to the laboratories of the University of Florence, Italy, where they have been bred repeatedly using standard methods (Chang & Peter, 1983). The first were raised in aquaria (100x40x50 cm) containing 160 liters of dechlorinated tap water and equipped with an adsorbing charcoal filter and aerator. They were maintained in darkness except during food administration and aquarium maintenance. Three to six times per week the fish were fed with frozen chironomid larvae. The water temperature was maintained constant at 29±1°C.

2.2 Cell lines

The CF-1 cell line was obtained from fin clips of adult cavefish adapting the protocol from (Vallone et al., 2007). Fin clips were treated with trypsin and plated in 50 mm cell culture dishes (Falcon) with Leibovitz´s L-15 medium (Gibco-BRL) supplemented with 20% fetal calf serum (Biochrom KG), 100 units/ml penicillin, 100 µg/ml streptomycin and 50 µg/ml gentamycin (Gibco BRL) and maintained in an atmospheric CO_2 and non-humidified cell culture incubator at 25°C. Once cells had grown to confluence, they were trypsinized and reseeded in flasks (Falcon). These fibroblast-like cells grow optimally as an adherent monolayer culture on normal tissue culture-treated plastic substrates. Cells were typically passaged once every two weeks by first using trypsin to induce detachment from the substrate followed by dilution in culture medium at a ratio of 1:2 and then seeding in fresh culture flasks. Typically, cells proliferated and returned to confluence within two weeks of passaging and confluent cultures could be maintained for up to one month without significant loss of viability.
A sub-line derived from the the zebrafish adult cell line AB9 (Kwok et al., 1998) was cultured under equivalent conditions. Cells were typically passaged once every two weeks

by first using trypsin to induce detachment from the substrate followed by dilution in culture medium at a ratio of 1:4 and then seeding in fresh culture flasks. Typically, cells proliferated and returned to confluence within one week of passaging and confluent cultures could be maintained for up to 1 month without significant loss of viability.

2.3 Genomic DNA extraction

A confluent flask (Falcon) of CF-1 cells or a section of cavefish *P. andruzzii* caudal fin was treated with 500 µl of lysis Buffer (10 mM Tris-HCl pH 8; 200 mM NaCl; 5 mM EDTA pH 8; 0,2% SDS) containing 7 µl of Proteinase K (20 mg/ml) after removing the culture medium and incubation for one hour at 55°C. Samples were centrifuged for 20 minutes at 12,000g at room temperature and the supernatants were transferred into new tubes and extracted once with phenol-chloroform. The aqueous phase was incubated with RNase A 50 µg/ml for 30 min at 37°C and then precipitated with 300 µl isopropanol, followed by centrifugation for 20 min at 12,000g at room temperature. The pellet was then washed with 500 µl 80% ethanol, air dried and resuspended in 100 µl DNase free water.

2.4 RNA extraction and reverse transcription

Total RNA was extracted from confluent CF-1 cell monolayers, AB9 cell monolayers or tissues from cavefish (*P. andruzzii*) and zebrafish (*D. rerio*) by lysing in TRIzol Reagent (Life Technologies). Addition of chloroform and subsequent centrifugation lead to phase separation under conditions where RNA remained water-soluble and proteins or DNA were partitioned in the lower, organic phase or at its interface. Total RNA was subsequently isolated from the aqueous phase by isopropanol precipitation followed by centrifugation and then rinsing the pellet using 75% ethanol.
Reverse transcription was performed with the extracted total RNA to produce cDNA using Superscript III RT (Invitrogen), according to the manufacturer's recommendations.

2.5 Cloning *Period2* sequence

In order to obtain the genomic sequence of cavefish *Period2* gene, genomic DNA was extracted as described in section 2.3.

The genomic structure of zebrafish *Per2* (ENSDARG00000034503) gene was used as a model to assess the putative number and position of exon-intron splicing sites in the homolog cavefish *Per2* sequence previously cloned from cDNA (full-length ORF, 3,777 bp). Primers flanking each of the exon-intron putative splicing sites were designed by Primer3 software.

By this approach, 13 of a total of 18 introns were amplified by using AmpliTaq DNA Polymerase (Applied Biosystems), bands of the predicted size were cloned into the pGEM-T easy Vector (Promega) and then sequenced (Microsynth). Subsequently, the DNAseq analysis performed on *P. andruzzii* confirmed the results obtained from this cloning approach, completing in addition the sequences of the remaining number of introns that were not possible to obtain before.

2.6 RT-PCR reactions

RT-PCR was performed with cDNA obtained as explained in section 2.4. All the primers were designed by Primer3 software and the sequences were amplified by using Go Taq DNA Polymerase (Promega). Reaction mixtures contained approximately 100 ng of template cDNA, 1 µl of each primer (each 10 µM), 10 µl of 5X GoTaq reaction buffer, 1 µl of dNTPs mix (each 10 mM), and 0,25 µl of Go Taq DNA polymerase (5 u/µl) in a total volume of 50 µl.

To validate the presence of the *Per2* transposon *in vivo*, primers flanking the transposon were designed on the basis of the sequence previously cloned from cavefish cDNA. The PCR amplification profile consisted of an initial denaturation step at 94°C for 2 min, followed by 30 cycles of 94°C for 30 sec, annealing at 63°C for 30 sec, elongation at 72°C for 1 min and a final extension at 72°C for 5 min.

For the *in vivo* validation of the intron retention in cavefish *Clock1b*, *Cry5, Cry4* and *Cry-dash* genes, primers detecting each intron were designed on the basis of the sequences obtained from the CF-1 cell line RNAseq. The PCR amplification profile consisted of an initial denaturation step at 94°C for 2 min, followed by 30 cycles of 94°C for 30 sec,

annealing for 30 sec at 60°C (*Clk1b*), 63°C (*Cry5*), 59°C (*Cry4* and *Cry-dash*), elongation at 72°C for 1 min and a final extension at 72°C for 5 min.

RT-PCR primers sequences:

Gene	**Primers sequences (from 5' to 3')**
CF *Per2* transposon	Forward CAGCTACAGCAGCACCATTGTACA Reverse GAATATGGCGTCTGGATGGGCTG
CF *Clk1b* intron	Forward TCACTAGCCAAGTGCCATGA Reverse CGGAGTTGAGGTGCAGGTAT
CF *Cry5* intron	Forward AGCTGCTGCTGGATTCTGAT Reverse TTTCGCATATGCTGCTTTCA
CF *Cry4* intron	Forward GCAAAGTCTGGAGGATCTGG Reverse CGCTTGAACCCCGAATAATA
CF *Cry-dash* intron	Forward CAAAAGGCACGTCCTGGTAT Reverse TTTTCCTGAAGGAGCAGATGA

2.7 Quantitative PCR (qPCR)

Quantitative PCR (qPCR) was performed with cDNA obtained as described in section 2.4. 4 µl of 1:10 diluted cDNA was pipetted in each well of a 96-well plate together with the SYBRgreen-Primer-MasterMix (Promega). qPCR was performed in an ABI StepOnePlus Real-Time PCR machine (Applied Biosystems) with a standard temperature cycle program, according to the manufacturer's conditions. PCR conditions were: 15 min at 95°C, then 40 cycles of 15 sec at 95°C, 30 sec at 60°C. The relative levels of each RNA were calculated by the $2^{-\Delta\Delta CT}$ method (CT standing for the cycle number at which the signal reaches the threshold of detection) (Livak & Schmittgen, 2001). The results obtained were subsequently standardized with *ß-actin* control data.

Each couple of primers was designed in a conservative nucleotide region between cavefish (*P. andruzzii*) and zebrafish (*D. rerio*), with the aim of cross-reacting with the sequences of both species. To do this, a pairwise sequence alignment by EMBOSS Needle software (EMBL-EBI) was performed between cavefish and zebrafish homolog sequences, in order to highlight high similarity shared regions and design primers to them.

The primers in cavefish were designed using the full-length ORF sequences of *Per2*, *Cry1a* and *ß-actin* formerly cloned and confirmed by RNAseq, while for all the opsins analyzed, the primers were designed based on exonic sequences obtained by DNAseq analysis. All zebrafish primers were designed using sequences obtained from the GenBank database.

qPCR primers sequences:

Gene	**Primers sequences (from 5' to 3')**
CF-ZF *Per2*	Forward CCGCAAAGTTTCCTTCGTCA Reverse CATTACTGCCCAGACTCCCA
CF-ZF *Cry1a*	Forward GGCTCCACGACAATCCTTCA Reverse TGGGGAAGACATCGGTAGGT
CF-ZF *Opn1lw1*	Forward ATGATGGCTCTGAGGTGTCC Reverse ATTCCCCATCACTCCAAGGG
CF-ZF *Opn1lw2*	Forward GCTATGCCTTCCACCCACTG Reverse GTGGACACCTCAGAGCCATC
CF-ZF *Opn1sw2*	Forward AAGGTGGCTGGTCATTTGCA Reverse GGAACCGCAAAGCAAAAGCA
CF-ZF *Opn4m3*	Forward ATTCTGCGGTGCTCTGTTTG Reverse CAGCCCATAGAGTAAGCCCA
CF-ZF *TMT-opsin*	Forward CTCATCTCTCTGGCTGTGCT Reverse GTCCAGTCTACAGAGCAGGT
CF-ZF *Opn3*	Forward CATGACTCTCTCGGGTCTGG Reverse GGTTCCATCCCAACAGAGGA
CF-ZF *Opn5*	Forward CCAGAGCTTTGTGATGTGCAT Reverse ACCACAGCATACGGGATCC
CF-ZF *ß-actin*	Forward GATGAGGAAATCGCTGCCCT Reverse GTCCTTCTGTCCCATGCCAA

2.8 Recording of adult locomotor activity

Cavefish locomotor activity was registered continuously by the use of infrared photocells (E3S-AD62, Omron) placed on the aquarium walls. The number of light-beam interruptions was counted and stored every 10 min by a computer connected to the

photocell. The analysis of locomotor activity records, representation of actograms and X^2 periodograms (Sokolove-Bushell test) were performed using the chronobiology software El Temps (version 1.228).

In light exposure experiments, adult cavefish were illuminated at a constant temperature of 29°C using one of the following light-sources:

• Monochromatic red-light-emitting diodes (LED, Kopa, 657nm)

• Monochromatic blue-light-emitting diodes (LED, Kopa, 468nm)

2.9 Deep Sequencing analysis

2.9.1 Stranded RNA sequencing library construction

Total RNA from *P. andruzzii* CF-1 cells and whole brain obtained from an individual adult fish was extracted as explained in section 2.4. RNA quality was checked using Agilent Bioanalyzer 2100 total RNA Nano chip (Agilent), and showed no sign of degradation (RNA index number > 9).

Sequencing libraries of stranded RNA were generated from 1 µg total RNA samples following the TruSeq RNA protocol v2 (Illumina) with the following modification: 1) at the 2^{nd} strand synthesis step 8 µl dUTP mix (NEB E7426G) and 4 µl of 2^{nd} strand synthesis Enzyme mix (NEB E7425G) were added to 25 µl of 1^{st} cDNA in final volume of 80 µl; 2) before PCR amplification 3 µl of USER enzyme (NEB M5505S) were added and samples incubated for 15 min in at 37°C.

2.9.2 Paired-end DNA sequencing construction

Genomic DNA was extracted from CF-1 freshly pelleted cells by adding 500 µl of 80 mM Tris pH 8.5, 200 mM NaCl, 0.5% SDS, 5 mM EDTA and 1 mg/ml Proteinase K (Sigma) and incubation at 65°C overnight. All subsequent steps were performed at room temperature unless otherwise specified. DNA was extracted by adding 0.5 ml Phenol/Chloroform/ IAA (25/24/1) (Ambion), gentle mixing for 10 min followed by centrifugation 10 minutes at 10,000g. DNA was precipitated by adding one volume of isopropyl alcohol 100% to the supernatant, followed by centrifugation for 10 min at 10,000g, washing with Ethanol 75% and centrifugation for 10 min at 10,000g. DNA pellet

was resuspended in 0.1 ml of water (Ambion) 10 min at 40°C. Quality of genomic DNA was assessed on 1% agarose gel; no sign of DNA degradation was observed.
For paired-end libraries, 1 µg of genomic DNA was sonicated with Covaris S220 under two different conditions in order to produce one partially overlapping library of averaged 180 bp fragments (Peak Incidence Power 175W, Duty factor 10%, cycles per burst 200, time 430 sec) and one non-overlapping library of averaged 380 bp fragments size (Peak Incidence Power 175W, Duty factor 10%, cycles per burst 200, time 360 sec). Paired-end libraries were constructed following the standard TruSeq DNA v2 protocol (Illumina), with an additional step of size selection at 290 bp +20% (fragment size plus size of flanking adapters) on a LabChipXT with DNA 750 kit (Perkin Elmer) before PCR amplification for the partially overlapping library.

2.9.3 Mate pair libraries construction

Mate pair libraries were prepared with the Nextera mate pair kit (Illumina) following the manufacturer's recommendations. One mate pair library was produced starting from 1 µg of DNA following the gel free method and four mate pair libraries were produced following the gel plus method. In the later case, 4 µg of tagmented sample were loaded on a 1% agarose gel, following recommendations, and bands cut under UV light (Eagle eye, Peqlab) to select DNA fragment at 20-17 kb, 17-10 kb, 10-7 kb and 7-3 kb.
Shearing of circularized DNA was performed with a Covaris S220 device (Peak Incidence Power 240W, Duty factor 20%, cycles per burst 200, time 40 sec).

2.9.4 Sequencing of libraries

Libraries size was validated on a Bioanalyzer with a DNA 1000 chip (Agilent) and quantified by qPCR (KAPA SYBR FAST ABI Prism, Peqlab) on a StepONE Plus device (ABI) following recommendations. Libraries concentrations were adjusted to 10 pM based on qPCR results. The five mate pair libraries were multiplexed to equal amount on a single lane of sequencing, while paired-end libraries from stranded RNA and the two from DNA were sequenced on individual lanes. Cluster production were made on the cBot using TruSeq PE Cluster Kit v3 – cBot™ - HS kit (Illumina). Paired-end reads (2x200 bp for the cavefish brain sample and 2x100 bp for all the other samples) were obtained on a Hiseq1000 using TrueSeq SBS v3 kits (Illumina).

Cluster detection and base calling were performed using RTAv1.13. De-multiplexing and quality of reads were assessed with CASAVA v1.8.1 and Fastx-tool kit v0.0.13.

2.9.5 *De novo* transcriptome assembly

Trimming of adapter sequences was performed with flexbarv2.4 using the parameters: *-n 18 --min-read-length 30 -f i1.8 --adapter-trim-end* ANY *--adapter-threshold 1 -ao 30*. Cleaned reads were assembled *de novo* with Trinity software (Grabherr et al., 2011), using the option --*SS_lib_type RF* for stranded RNAseq data.

Contigs with a minimal size of 200 bp where used. Redundant sequences with high similarities were collapsed to single contigs with cd-hit-est v4.6.1 (2012-08-27) with the option *-c 0.95*. Mapping of the reads back to the contigs was performed with Bowtie v0.12.7 via the script *alignReads.pl* provided in Trinity package. Prediction of CDS contigs was performed with *transcripts_to_best_scoring_ORFs.pl* (TransDecoder utility, within Trinity package) with the *-S* option for strand specific data.

Homology search against zebrafish Refseq nucleotide sequences as well as homology search against zebrafish and Uniprot Knowledgebase (UniProtKB, released 18-09-2013) Refseq proteins was performed respectively with Blastn and Blastx, with an E-value cutoff of 10^{-10}. Best hits were selected based on best E-value.

For functional analysis of transcripts, Uniprot IDs obtained by Blastx were incorporated along annotation files of EuKaryote clusters of Orthologous Groups (KOG) (eggNOG v3.0, Powell et al., 2011) in a MysQL relational database.

2.9.6 *De novo* genome assembly

Trimming of the 5 mate pairs reads was performed with Cutadapt-1.0 with the options *–a* GATCGGAAGAGCACACGTCTGAACTCCAGTCAC *-m 20* in order to remove the 3' sequence of the reads that match the adapter, followed directly by an additional trimming of the junction adapter with the option *-b* CTGTCTCTTATACACATCT *-b* AGATGTGTATAAGAGACAG *-m 20*. The reads from the two fragment libraries were trimmed with Flexbar_v2.4 with the options *-n 18 -z* GZ *-r read1.fastq -p read2.fastq -j --min-read-length 30 -f i1.8 --adapter-trim-end* ANY *--adapter-threshold 1 -ao 30* and providing the FASTA sequences of the following adapters:

AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC;

AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTA

TCAT;GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT;
ATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGA
TCT.

De novo genome assembly was performed with ALLPATHS-LG (Gnerre et al., 2010, version #47539) on a high performance AMD-Opteron-6100-Magny-Cours-G34Quad-12-Cores Linux workstation with 256 Gbytes RAM.

Inserts size were confirmed during the assembly with ALLPATHS-LG. Zebrafish Refseq transcripts and Trinity transcripts were mapped on scaffolds with Blat, through the option *minIdentity=10*, and then parsed to get single hits with highest number of matches.

III. Results

3.1 Cavefish *Period2* analysis

In order to understand the origin of the abnormal infradian clock phenotype in the cavefish *P. andruzzii*, as well as the origin of the loss of light entrainment and the reduced temperature compensation, we hypothesized that this might reflect mutations affecting the cavefish's clock gene repertoire. To date, the most relevant mutation encountered in a cavefish core clock gene has been the insertion of a transposable element and a premature stop codon mutation in the *Period2* gene, which leads to the deletion of the C-terminal region of the coding sequence. This portion of the PER protein has been implicated in the interaction with several key clock proteins such as the cryptochromes (CRYs). Interaction with CRY is an essential step for the molecular feedback loop of the core circadian clock (Tomita et al., 2010). In the case of cavefish the prediction would be that, since this region absent, then also functionality of PER proteins within the core clock mechanism may also be abnormal. Thus, by studying the relevance of this type of mutation, we can exploit evolution as a tool to explore the function of normal clock proteins and their regulatory networks. For this reason, we have chosen to investigate this core clock gene in detail.

3.1.1 Cloning cavefish *Period2* gene

As a first step, the full Open Reading Frame (ORF) of cavefish *Per2* gene was cloned from CF-1 cell line cDNA. Our cloning strategy involved many different steps due to the large size of the ORF and the low expression level of *Per2* in the CF-1 cell line. The cavefish *Per2* cDNA has a 3,777 bp ORF and encodes a protein of 1,259 amino acids. As in all other *Period* gene homologs, the N-terminal region contains PAS and the PAC motifs (PAS A, aa 247-312; PAS B, aa 387- 451; PAC, aa 460-501) followed by a nuclear export signal (NES, aa 525-535), and the phosphorylation target site for casein kinase 1 epsilon (CK1e, aa 747-760), which is involved in regulating the degradation of this protein. The alignment of cavefish PER2 PAS A, PAS B and PAC domains with homologous sequences is shown in Figure 7.

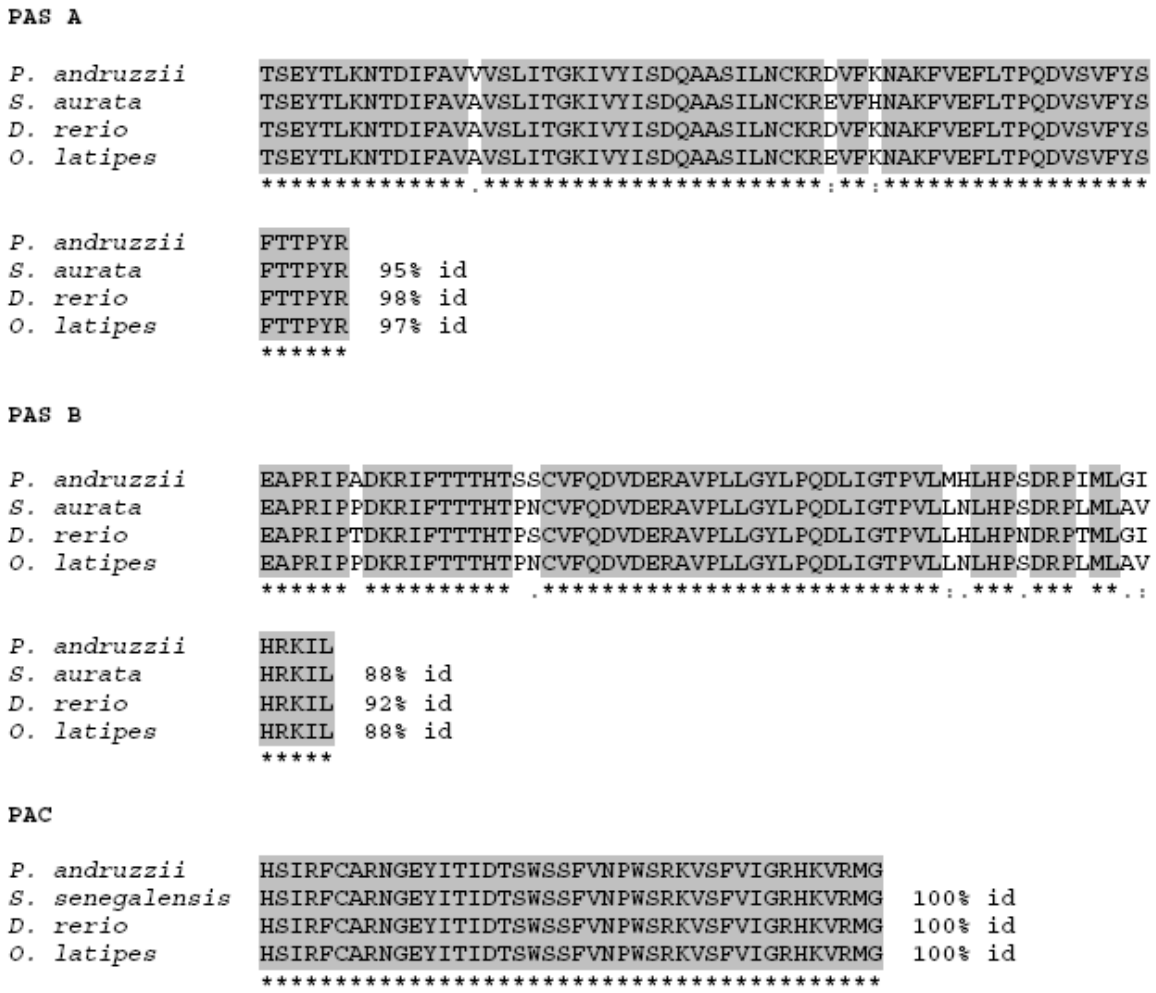

Figure 7. Alignment of cavefish PER2 PAS A, PAS B and PAC domains. PAS A, PAS B and PAC are fundamental domains which can mediate homodimeric and heterodimeric PER-PER interactions, as well as interactions with transcription factors and kinases. These domains have been aligned between *P. andruzzii* and the homologous sequences from *Sparus aurata* (AGC92272.1), *Danio rerio* (AAI63549.1) and *Oryzias latipes* (XP_011481351.1). The PAC domain has not been characterized yet in the partial sequence of *S. aurata Per2* gene present in GenBank, which has been replaced for this reason in the alignment with the homologue sequence from *Solea senegalensis* (CAQ86911.1). The percentages of identity between *P. andruzzii* and each of the homologous sequences are indicated on the right side of the alignments. Multiple sequennce alignment performed by Clustal Omega (EMBL-EBI), free online software.

In the 3´region of the ORF, at position +2,836 from the transcription start site, we encountered a large insertion of 225 bp, similar to a reverse transcriptase coding sequence (transposon). The insertion preserves the transcript's open reading frame and inserts 75 amino acids in the protein sequence (transposon, aa 856-931). Downstream of this insertion, a single nucleotide substitution mutation leads to a premature stop codon at aa 1,259. Due to the premature stop codon, the CRY binding domain, that is located in

zebrafish at position aa 1,315-1,334, is lost in cavefish. The *Per2* sequence contains in addition a predicted nuclear localisation site (NLS). In zebrafish the NLS site is at position aa 877-892 but in cavefish, as a result of the transposon insertion, the predicted NLS is shifted to position aa 954-969, between the transposon and the premature stop codon (Figure 8A). The transposable element detected in cavefish *Per2* ORF shows some similarity with other homologous sequences derived from genomic libraries of other species of fish. The sequence of the cavefish transposon and its alignment with some of these homologous sequences, selected by the lowest Blast E-value, is shown in Figure 8B.

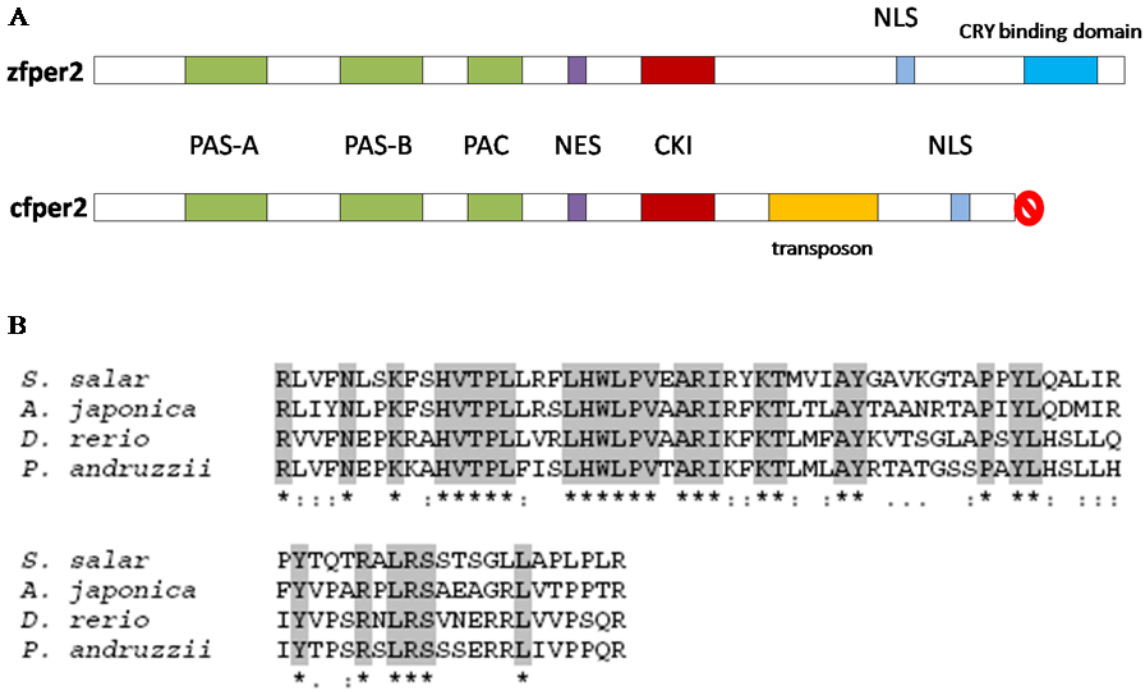

Figure 8. Cavefish and zebrafish *Period2* ORF and transposon sequence alignments. (A): Both genes contain PAS-A, PAS-B and a PAC domain. The nuclear export site (NES) and the target site for casein kinase 1 epsilon (CKI) are also present, as well as a nuclear localisation site (NLS). The cf *Per2* carries a transposable element of 225 bp and a premature stop codon mutation, whereas the zf *Per2* does not show these mutations and contains at the 3´ terminus a normal CRY binding domain. (B): The transposable element from cavefish shows some similarity with other homologous sequences derived from genomic libraries of different species. In particular, in the figure is shown the alignment between *Phreatichthys andruzzii* and *Salmo salar* (56% of identity), *Anguilla japonica* (56%) and *Danio rerio* (72%). GenBank accession numbers of cf homologous sequences: *D. rerio* (BAE46430.1), *A. japonica* (BAD72127.1) and *S. salar* (ABQ01988.1). Multiple sequennce alignment performed by Clustal Omega (EMBL-EBI), free online software.

Following our cloning of the full ORF of cavefish *Per2*, we proceeded to clone the introns sequences from the genomic DNA in order to generate a complete map of the gene. We used the genomic structure of zebrafish *Per2* as a guide to estimate the number, size and position of exon-intron splicing sites in the cavefish *Per2* gene, and we thereby designed PCR primers flanking each of the exon-intron putative splicing sites. Using these primers, we were able to amplify a set of 13 out of a total of 18 introns from the cavefish genomic locus. The length of the cloned introns and their order, from the 5' to the 3' of the *Per2* gene, are the following: intron 3 (670 bp), intron 4 (1,271 bp), intron 5 (113 bp), intron 6 (446 bp), intron 7 (2,701 bp), intron 8 (1,066 bp), intron 9 (195 bp), intron 10 (173 bp), intron 11 (1,081 bp), intron 12 (127 bp), intron 13 (120 bp), intron 14 (1,249 bp), intron 15 (574 bp).

Subsequently, the Deep Sequencing analysis performed on *P. andruzzii* genomic DNA (DNAseq analysis, see Result section 3.4), confirmed the results obtained from this initial characterization. The sequences of the remaining introns that were not possible to amplify previously, were obtained from this DNAseq analysis. These intron sequences contain some sections with undetermined nucleotides (polyN) due to gaps in the genome assembly, but the length of each intron could be still reconstructed: intron 1 (13,461 bp), intron 2 (16,496 bp), intron 16 (1,271 bp), intron 17 (4,868 bp), intron 18 (3,290 bp).

The complete map of cavefish *Per2* genomic sequence is shown in Figure 9.

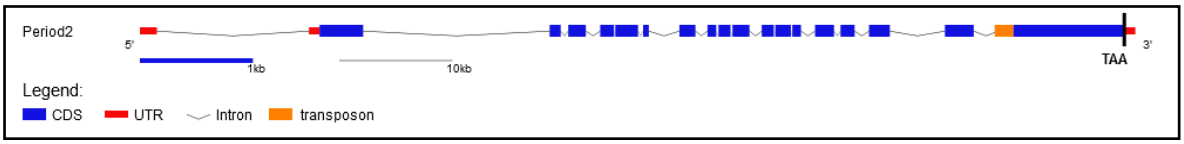

Figure 9. Schematic structure of the cavefish *Period2* genomic sequence. The cavefish *Per2* gene is composed by a total of 18 exons (in blue), separated by introns (grey lines). A 5' and 3' UTRs are present (in red). In the figure, the intron spacing is rescaled (1:10) due to the excessive length of the introns, while the exons are to scale (1:1). The transposon position is marked in yellow and the premature stop codon is indicated (TAA). GSDS software was used to built the figure (Guo et al., 2007).

3.1.2 Transposon validation and expression levels of cavefish *Per2*/*Cry1a*

The insertion of a transposable element in the *P. andruzzii Period2* gene preserves the ORF of the cDNA sequence and inserts 75 aa at the C-terminus of the protein. This fact can lead to considerable changes in the three-dimensional structure of the PER2 protein, possibly altering its stability as well as its ability to interact with other components of the oscillator. It could thereby potentially have a significant impact on the entire clock machinery.

We initially wished to confirm the presence of the transposon in cDNA prepared from various adult tissues. For this purpose, we designed primers flanking the transposon based on the previously characterized ORF sequence and we used them for RT-PCR on cDNA obtained from cavefish brain (central clock oscillator) and fin (peripheral clock). In the brain, two clear bands were obtained, pointing to the existence of two different isoforms of cavefish *Per2*: one band of ~912 bp, corresponding to the expected length of the isoform containing the transposon insertion, and one band of ~692 bp, corresponding to an isoform in which the transposon is absent (Figure 10). These predictions were confirmed by subsequent sequencing of the amplified products.

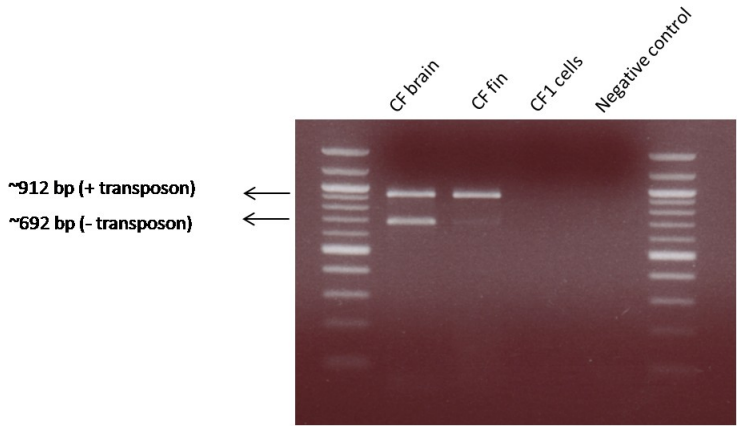

Figure 10. RT-PCR analysis of the cavefish *Per2* transposon "*in vivo*". Primers flanking the transposon insertion were designed and tested on cavefish brain and fin cDNA. Two clear bands were obtained, corresponding in length to two different isoforms of the gene (+/- transposon). There is a differential intensity of the bands between the two tissues, highlighting distinct levels of expression. A negative control sample without cDNA was included in the PCR reaction and shows no sign of contamination in the reagents.

In the case of the brain RT-PCR reaction, from the intensity of the two bands, the two isoforms seem to have the same level of expression. Also in the fin these two bands were

amplified, but the intensity of the band corresponding to the isoform without the transposable element (~692 bp) was considerably lower than the one corresponding to the isoform that carries the transposon (~912 bp).

Abberant transcript processing and the presence of premature stop codons have in several cases being linked with increased transcript turnover and thereby reduction in the levels of the mRNA (Braunschweig et al., 2014). For this reason, we next wished to compare the relative levels of *Per2* expression in cavefish and zebrafish as a control. In parallel, we also analysed the relative expression levels of *Cry1a*, another key clock gene but which does not show any significant mutations in *P. andruzzii*. Both genes serve as negative transcriptional regulators at the core of the clock mechanism, as well as showing light inducible expression. For this reason in the case of zebrafish, they are predicted to participate in the resetting of the clock by light.

Per2 and *Cry1a* qPCR primers were designed in conservative nucleotide regions showing high-shared similarity between cavefish and zebrafish, with the aim of cross hybridization with the transcripts of both species. We tested the expression levels of *Per2* and *Cry1a* in cavefish and zebrafish brain, the site of a central clock oscillator, plus the CF-1 cell line and the equivalent AB9 adult fin-derived cell line in zebrafish, as "*in vitro*" systems models. The results are shown in Figure 11.

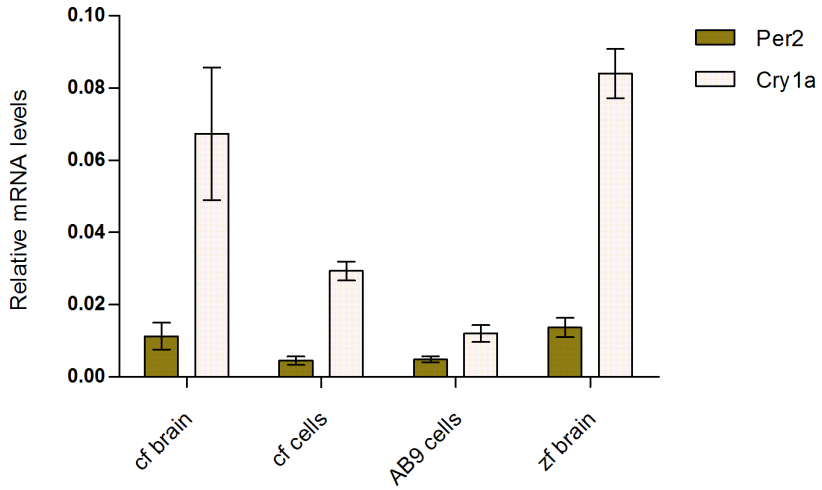

Figure 11. *Per2* and *Cry1a* expression levels in cavefish and zebrafish tissues. qPCR analysis of *Per2* and *Cry1a* in cavefish brain, CF-1 cells, AB9 cells and zebrafish brain. The pattern of expression is similar between the two species, with higher levels of expression in the brain than in the cell line. Within each tissue, a significantly higher expression of *Cry1a* compared to *Per2* is observed. Mean and SEM are plotted. The graph was plotted with GraphPad Prism 5.0 (GraphPad Software Inc.).

The pattern of expression is similar between the two species, showing a general higher level of gene expression in the brain compared to the cell line. Comparing *Per2* and *Cry1a* within each tissue, we observed a significant difference in the expression of these genes, with *Cry1a* more highly expressed than *Per2* in all the four cells and tissues examined.

3.1.3 Next Generation Sequencing: a new global approach

Does the *Per2* gene mutation represent the only example of clock gene disruption in the cavefish? In the last few years our group has performed a preliminary molecular characterization of the *P. andruzzii* circadian clock; clock genes have been identified and cloned from *P. andruzzii* cDNA based on their sequence homology with zebrafish (Cavallari et al., 2011). However, many clock gene sequences were incomplete and so we were unable to assess the extent of cavefish mutations. For this reason, we have decided to adopt a "global" strategy of investigation in *P. andruzzii* by transcriptome analysis (RNAseq) and genome sequencing (DNAseq). Two *de novo* transcriptome assemblies were produced from two different *P. andruzzii* RNA libraries: one assembly was generated from the CF-1 cell line RNA and one from the cavefish whole brain RNA obtained from an individual adult fish. One *de novo* genome assembly was generated from *P. andruzzii* genomic DNA libraries. All the technical details about RNAseq and DNAseq workflows (HiSeq 1000, Illumina) are explained in Materials and Methods, section 2.9.

3.2 RNAseq analysis in *Phreatichthys andruzzii*

3.2.1 General statistics on *Phreatichthys andruzzii* RNAseq

A total number of 344 million reads (2x100 bp) was generated from CF-1 cell line RNAseq, corresponding to a size of 34 giga base pairs (Gb), while from the cavefish brain RNAseq a total number of 553 million reads (2x200 bp) was generated, corresponding to a size of 104 Gb. Cleaned reads were assembled *de novo* using Trinity software (Grabherr et al., 2011) and homology search against Refseq nucleotide sequences from *D. rerio* was performed with Blastn. The statistics concerning the transcriptome assemblies are shown in Table 1.

The brain assembly is composed of a larger number of contigs compared to the CF-1 cell line but more fragmented, showing a lower N50 (i.e. the contig size above which 50% of the total length of the sequence assembly can be found) and a lower average contig length. This is most probably an indication of the great complexity and variety, in length and in levels of expression, of the transcripts from this tissue. Nucleotide blast (Blastn) showed that the longest contig in the CF-1 cell line corresponds to *microtubule-actin crosslinking factor 1* gene, while in the brain the longest contig corresponds to the *titin a* gene. The raw sequence coverage, calculated on the basis of the zebrafish transcriptome, is estimated respectively ~42 X for the CF-1 cell line and ~130 X for the brain.

	CF-1 cell line	**Brain**
Number of contigs	**272,717**	**380,678**
N50 (bp)	**2,187**	**1,635**
Average contigs length (bp)	**1,065**	**800.6**
Longest contig (bp)	**24,836**	**35,275**

Table 1. *P. andruzzii* CF-1 cell line and brain *de novo* transcriptome assemblies. Trinity software (Broad Institue, MIT) was used to assemble the cleaned reads, with the option for stranded RNAseq data. Redundant sequences with high similarity were collapsed into single contigs and only the contigs with a minimal size of 200 bp where considered.

Subsequently, by the use of the TransDecoder utility included in the Trinity package, we proceeded to identify all the candidate coding sequences within the transcripts generated by *de novo* assembly. Translated nucleotide sequences of all the potential ORFs found have been blasted against Uniprot database (*Homo sapiens*, *Danio rerio, Astyanax mexicanus* and *Oryzias latipes*) for homology search and annotation (Blastx), and best hits were selected, based on best E-value. By this procedure, we obtained a total of 56,462 unique Uniprot identifiers (IDs) for the CF-1 cell line, and a total of 35,441 unique Uniprot IDs for the brain. The next step was to utilize the Uniprot IDs previously identified for the functional analysis of the transcripts. To do this, the list of all the IDs obtained from the CF-1 cell line and brain has been incorporated in the eggNOG database (eggNOG v3.0, Powell et al., 2011), which provides the assembly of proteins into orthologous groups and their functional annotation, using an automated procedure based on the EuKaryotic Orthologous Groups (KOG) approach (Tatusov et al., 2003). The orthologous groups are in this way annotated with functional categories and identified by letters of the alphabet

according to the KOG classification. At the end of this process, a total of 7,167 proteins for the CF-1 cell line and 10,634 for the brain have been functionally annotated. The abundance, in percentage, of each KOG functional category for the CF-1 cell line and the cavefish brain annotated proteins is shown in Figure 12. As can be seen, the distribution pattern of the functional clusters is very similar between CF-1 cell line and brain, marking an interesting point for the comparability of "*in vitro"* and "*in vivo"* systems.

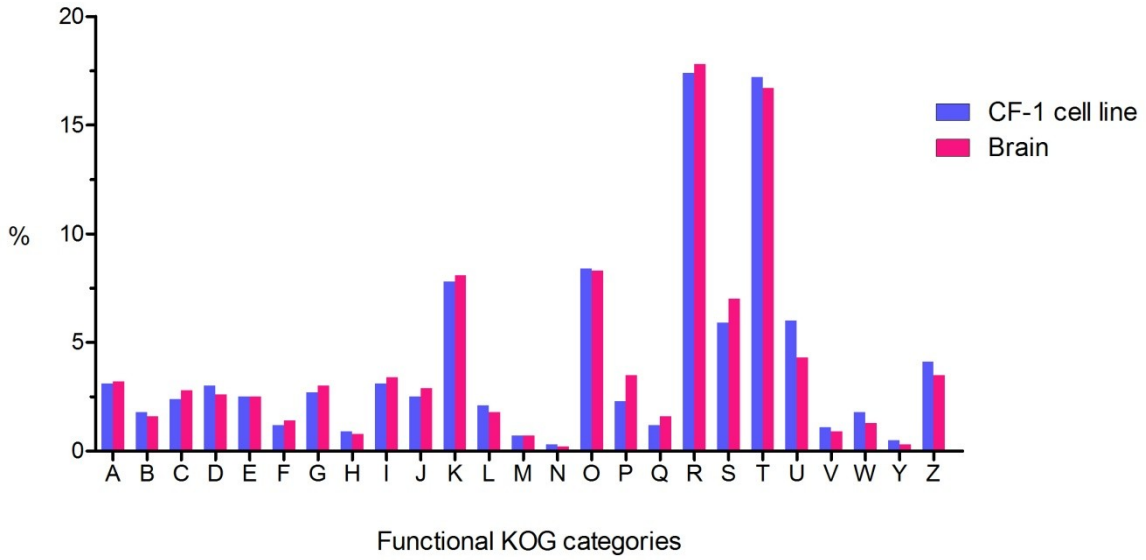

Figure 12. Functional analysis of *P. andruzzii* CF-1 cell line and brain Uniprot IDs, classified by eukaryotic clusters of orthologs (KOGs). Classification of functional KOG categories: (A) RNA processing and modification, (B) chromatin structure and dynamics, (C) energy production and conversion, (D) cell cycle control and mitosis, (E) amino acid metabolism and transport, (F) nucleotide metabolism and transport, (G) carbohydrate metabolism and transport, (H) coenzyme metabolism, (I) lipid metabolism, (J) translation, (K) transcription, (L) replication and repair, (M) cell wall/membrane/envelope biogenesis, (N) cell motility, (O) post-translational modification, protein turnover, chaperone functions, (P) inorganic ion transport and metabolism, (Q) secondary metabolites biosynthesis, transport and catabolism, (T) signal transduction, (U) intracellular trafficking and secretion, (Y) nuclear structure, (Z) cytoskeleton, (R) general functional prediction only (typically, prediction of biochemical activity), (S) function unknown. The graph was plotted with GraphPad Prism 5.0 (GraphPad Software Inc.).

3.2.2 The characterization of the cavefish circadian oscillator from RNAseq

As already mentioned, one of the major goals of our Deep Sequencing analysis in *P. andruzzii* was to complete and validate our characterization of clock gene sequences, as well as the genes involved in photoreception and the light input pathway. This would provide us with a global and comprehensive overview of the mutations affecting the clock

machinery and thereby enable us to explore the causes of the "aberrant" clock phenotype that characterizes this species of cavefish. In Table 2 are reported all the full-length sequences of clock genes and visual/non-visual opsins that we obtained by our RNAseq analysis in the cavefish CF-1 cell line and brain, and a comparison, in bp length, with the *D. rerio* homologous sequences from GenBank.

CLOCK GENES

Gene	Obtained from	*P. andruzzii* (linear RNA)	*D. rerio* (linear RNA)
Per1a	CF-1 cell line/brain RNAseq	6,074 bp	6,172 bp
Per1b	CF-1 cell line/brain RNAseq	6,084 bp	6,818 bp
Per3	CF-1 cell line/brain RNAseq	4,676 bp	4,809 bp
Cry1a	CF-1 cell line/brain RNAseq	3,359 bp	3,361 bp
Cry1b	CF-1 cell line RNAseq	3,062 bp	3,006 bp
Cry2a	CF-1 cell line RNAseq	4,356 bp	4,524 bp
Cry3	CF-1 cell line/brain RNAseq	3,775 bp	2,069 bp
Cry5	CF-1 cell line RNAseq	1,960 bp	2,363 bp
Clock1a	brain RNAseq	5,247 bp	2,682 bp
Bmal1a	CF-1 cell line/brain RNAseq	4,018 bp	2,343 bp
Bmal2	CF-1 cell line/brain RNAseq	4,155 bp	7,462 bp
Rorα	CF-1 cell line/brain RNAseq	7,223 bp	1,779 bp

VISUAL/NON-VISUAL OPSINS

Gene	Obtained from	*P. andruzzii* (linear RNA)	*D. rerio* (linear RNA)
Opn1sw2	brain RNAseq	1,435 bp	1,425 bp
Opn4m2	brain RNAseq	2,064 bp	1,668 bp
Opn4x1	CF-1 cell line RNAseq	2,229 bp	1,383bp
Opn5	brain RNAseq	1,824 bp	1,206 bp
Valopa	brain RNAseq	1,919 bp	1,869 bp
Rgra	CF-1 cell line/brain RNAseq	2,095 bp	1,107 bp

Table 2. Clock gene and visual/non-visual opsin gene full-length sequences obtained by RNAseq analysis in *P. andruzzii.* The genes listed on the table are considered as full-length coding sequences (CDS) plus 5' and 3' untranslated regions (UTRs). The sequences obtained from cavefish RNAseq were compared with the homologous sequences from *D. rerio* (Blastn best hits) present in GenBank. Many other clock genes and visual/non visual opsins for which we have obtained only partial sequences, as well as previously cloned genes, are not indicated. (Cavallari et al., 2011).

Genes expressed at a higher level in either CF cells or the brain (i.e. with a high number of raw reads and consequently a higher coverage) were more easily assembled as full-length contigs than the sequences of genes expressed at a lower level (i.e. with a low number of raw reads), whose assembly provided only partial fragments.

This new RNAseq analysis in cavefish confirmed the single base substitution mutations leading to premature stop codons in the melanopsins *Opn4m2* and *TMT-opsin*. The mutations in these two key non-visual opsins involved in peripheral circadian photoreception have been previously cloned and functionally tested in detail by our group (Cavallari et al., 2011). A full-length sequence of *Opn4m2*, shown in the Table 2, was obtained from our cavefish brain RNAseq analysis, while for the *TMT-opsin* we obtained only partial fragments from both transcriptomes due to very low expression levels. Another key issue was to confirm by RNAseq the transposon insertion and the premature stop codon in the C-terminal region of the cavefish *Per2* gene. Unfortunately, in this case, we obtained only partial fragments of the *Per2* coding sequence from both the CF-1 cell line and brain transcriptomes. These fragments cover the 5' region of the *Per2* coding sequence, and so it was not possible to further validate the presence of these mutations, which are located near the 3' end of the sequence.

During our characterization of clock and photoreceptor genes by RNAseq, we detected a substantial enrichment of transcripts carrying unspliced introns, a finding that was not revealed by our previous "manual" cloning and sequencing approach. Given the potential importance of this discovery for understanding the basis of the aberrant clock phenotype in this species, we chose to investigate it in more detail.

3.3 Intron retention in cavefish mRNA

The elevated intron retention in cavefish mRNA was revealed during the characterization of the clock genes and visual/non-visual opsin sequences obtained by RNAseq from the CF-1 cell line and cavefish brain. Each gene is represented in the transcriptome database by a set of contigs, either partial or full-length. In order to compare the length and the similarity of the cavefish clock genes and visual/non-visual opsins sequences with the zebrafish homologs, we performed a one-by-one pairwise sequence alignment (EMBOSS Needle, free online software) of each single contig with the homologues sequences from *D. rerio* present in GenBank. By this procedure, we detected different structures and proportions of the contigs related to each gene. More precisely, for each gene, some of the

contigs that we found were considered “normal”, meaning that the coding sequence is not interrupted by any insertions and shows the highest similarity with its zebrafish counterpart; these contigs have been taken as a model for the gene description (e.g. the genes listed in Table 2). However, in addition, for many genes other contigs were encountered, corresponding to sequences containing one or more unspliced intron(s). The frequent consequence of these retained introns, is to introduce a premature stop codon in the coding sequence which leads to a prematurely truncated protein.

To determine the position and precise nature of these intron insertions, each intron-containing contig was aligned with the zebrafish homologous genomic sequence present in the Ensembl database, in order to compare the intron-exon junction positions. By this approach we were able to confirm by homology the presence of *bona fide* introns. The same contigs were then translated into the relative amino acid sequences (ExPASy Translate, free online software), to test whether the intron insertions indeed generated premature truncated coding sequences.

We were able to classify the contigs containing retained introns into three types, based on the position of the intron in the coding sequence (intron flanked by exons or intron at the 5’ or 3’ region of the sequence, Figure 13).

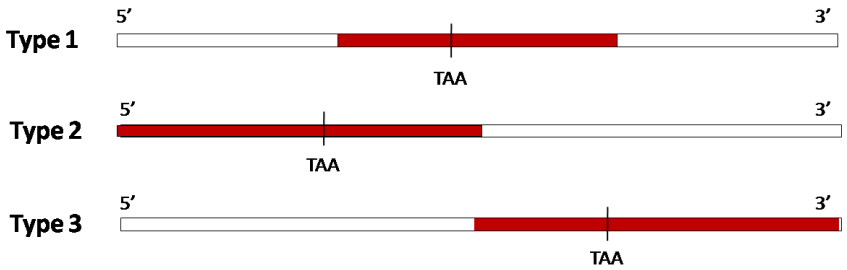

Figure 13. Structures of the contigs containing retained introns in cavefish mRNA. In contigs with retained introns found in cavefish mRNA, the intron (red bar) position could be detected either inside the coding sequence of the gene (Type 1, intron flanked by exons), at the 5’ region (Type 2, contigs starting with an intron) or at the 3’ region (Type 3, contigs ending with an intron). The premature stop codons introduced by the introns are indicated with TAA, as an example.

Considering the observed frequency of intron retaining transcripts, we defined three categories of genes: genes whose contigs are correctly spliced *in toto*, genes whose contigs are affected by intron insertions *in toto*, and genes whose contigs are composed of a mixture of spliced and unspliced transcripts. The percentage of these three transcript categories encountered in the four main types of clock genes (*Per*, *Cry*, *Clk* and *Bmal*)

within the CF-1 cell line and brain transcriptomes are summarized in Figure 14. As a control, we analysed the zebrafish transcriptome prepared from 24 hpf embryos by one of our collaborating groups at the Karlsruhe Institute of Technology (KIT) in Germany. The proportion of the three different transcript categories are very similar when comparing the CF-1 cell line and brain transcriptomes, however there are major differences compared to zebrafish. The proportion of correctly spliced clock genes in zebrafish is 78%, in contrast to much lower levels in the CF-1 cell line (22%) and cavefish brain (17%). Interestingly, while there are no zebrafish genes where all transcripts are affected by intron insertions, 22% of genes show contigs both with and without retained introns, demonstrating that a basal low level of intron retention is also present in zebrafish clock genes.

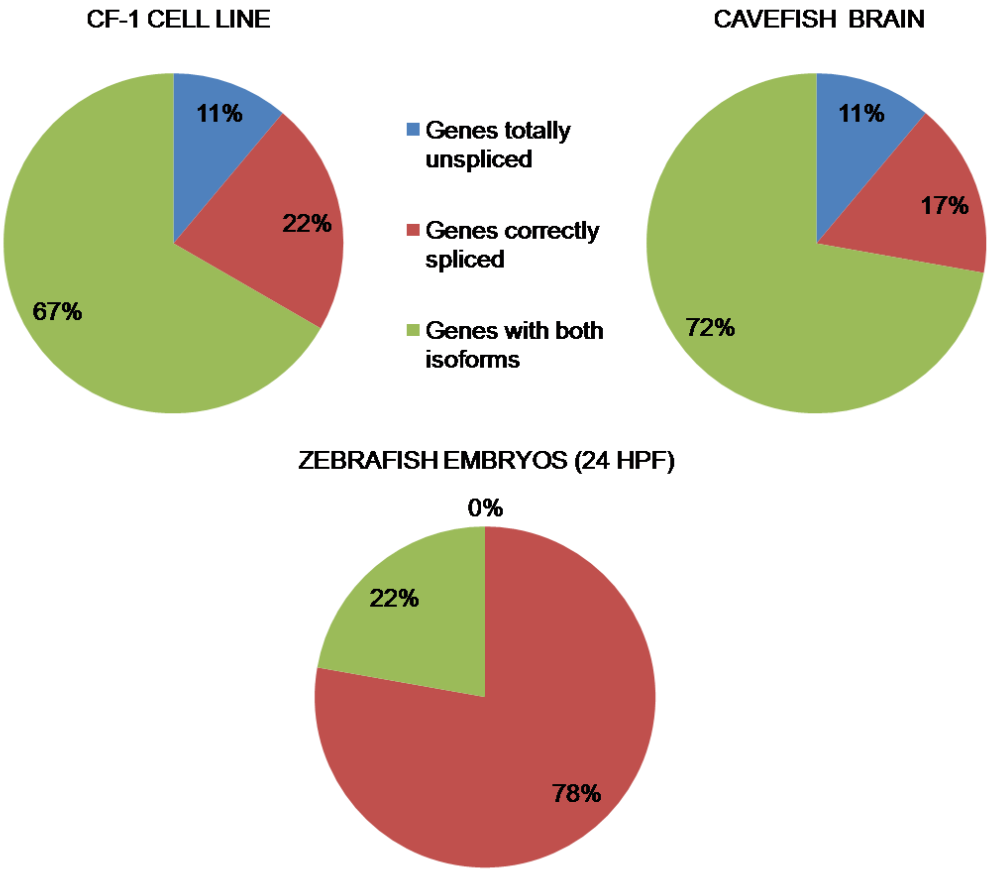

Figure 14. Different proportions of intron retention in core clock genes families (*Per*, *Cry*, *Clk* and *Bmal*), from CF-1 cell line, cavefish brain and zebrafish embryos (24 hpf) transcriptomes.

The percentages of spliced/unspliced contigs have been calculated considering all the representative transcripts (full-length and partial) of 18 genes, belonging to the four types of core circadian clock genes.. In detail: *Period* family (*Per1a*, *Per1b*, *Per2* and *Per3*), *Cryptochrome* family (*Cry1a*, *Cry1b*, *Cry2a*, *Cry2b*, *Cry3*, *Cry4*, *Cry5* and *Cry-dash*), *Clock* family (*Clk1a*, *Clk2* and *Clk1b*) and *Bmal* family (*Bmal1a*, *Bmal2* and *Bmal1b*). The zebrafish transcriptome database derived from 24 hpf embryos has been provided by courtesy of our collaborator Dr. Olivier Armant, NGS Core Facility, KIT, Germany.

The situation concerning the visual and non-visual opsin gene transcripts seems to be quite different compared to the clock genes. In Figure 15 is shown the results of a comparison between the visual and non-visual opsins expressed in the CF-1 cell line and cavefish brain transcriptomes, as well as in the zebrafish transcriptome from 24 hpf embryos. Relatively few of these gene sequences were obtained as full-length CDSs by RNAseq in cavefish (see Table 2), with the majority represented by partial fragments. This is most probably the reflection of the relatively low expression levels of these transcripts. Similarly, in the zebrafish 24hpf embryo transcriptome, only a few number of them has been detected possibly because of extremely low transcript levels in whole body RNA extracts prepared from embryos. One of the future objectives is of course to compare the transcriptome data from cavefish, especially that from the adult brain, with a more comparable transcriptome dataset from adult zebrafish. Interestingly, there is an almost complete lack of intron retention in the contigs representing visual and non-visual opsins. Only in two contigs, corresponding to *Opn1mw3*, obtained from cavefish brain, and *Rrh*, obtained from CF-1 cell line, introns insertions have been found. All the other contigs appear to be normally spliced, although this conclusion is based on short, incomplete transcript contigs.

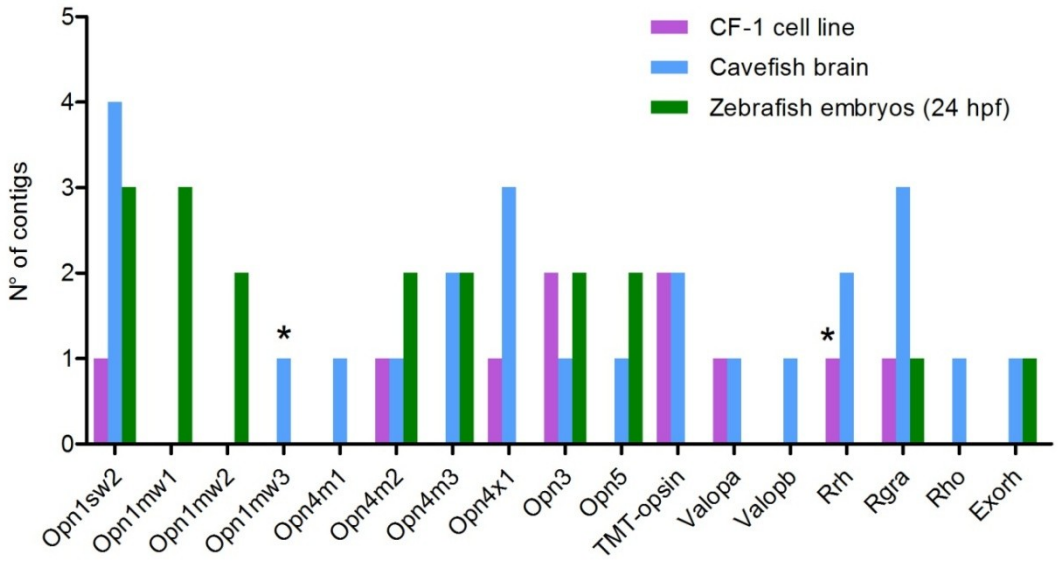

Figure 15. Number of contigs corresponding to each visual/non-visual opsin expressed in CF-1 cell line, cavefish brain and zebrafish embryos (24 hpf) transcriptomes. The highest number of contigs for each gene, as well as the greatest variety of opsin genes expressed is present in the cavefish adult brain transcriptome, followed by the zebrafish embryos at 24 hpf and CF-1 cell line transcriptomes. The only two contigs that have been found containing retained introns are marked with asterisks. The graph was plotted with GraphPad Prism 5.0 (GraphPad Software Inc.).

In order to validate the phenomenon of intron retention, an *in vivo* assay for the presence of intron sequences in clock gene transcripts was required. Thus, we first selected a subset of clock gene contigs in which retained introns were detected by our bioinformatic analysis, and which were also representative of the three different types of sequence structures based on the position of the introns (Type 1, 2 and 3, see Figure 13). A set of primers was designed in order to detect and amplify the intron sequences in each of the selected genes, using RT-PCR performed on RNA from different tissues of an adult cavefish, plus the CF-1 cell line. A number of intron insertions were confirmed by this approach (shown in Figure 16).

Specifically, *Clk1b* and *Cry5* were selected as examples of Type 1 retained introns (intron flanked by exons). In both cavefish transcriptomes (CF-1 cell line and brain), they were also found to belong to the category of genes whose contigs are a mixture of normal and intron-retaining transcripts. With the exception of *Clk1b* in cavefish heart, two bands were amplified in each of the cavefish tissues. These two bands correspond to the predicted length of the two isoforms, with and without retained introns. Differences in the relative intensity of the two bands in different tissues would seem to indicate that the level of intron retention is tissue or cell type specific.

Cry4 and *Cry-dash* are two important genes belonging to the *Cryptochrome* family. *Cry4* is a gene that has been speculated to be involved in photoreception and shares high homology with *Drosophila Cry* while *Cry-dash*, as well as *Cry5*, are involved in the repair of UV-light damaged DNA. In both cavefish transcriptomes, both of these genes belonged to the category whose contigs are affected by retained introns *in toto*, and thereby encode prematurely terminated proteins. The sequences shown in Figure 16 are examples of Type 2 and Type 3 retained introns (intron at the 5' or at the 3' end of the sequence). Bands of the predicted length were amplified in each of the cavefish tissues, confirming also in this case the presence of the retained introns deduced by the Deep Sequencing analysis.

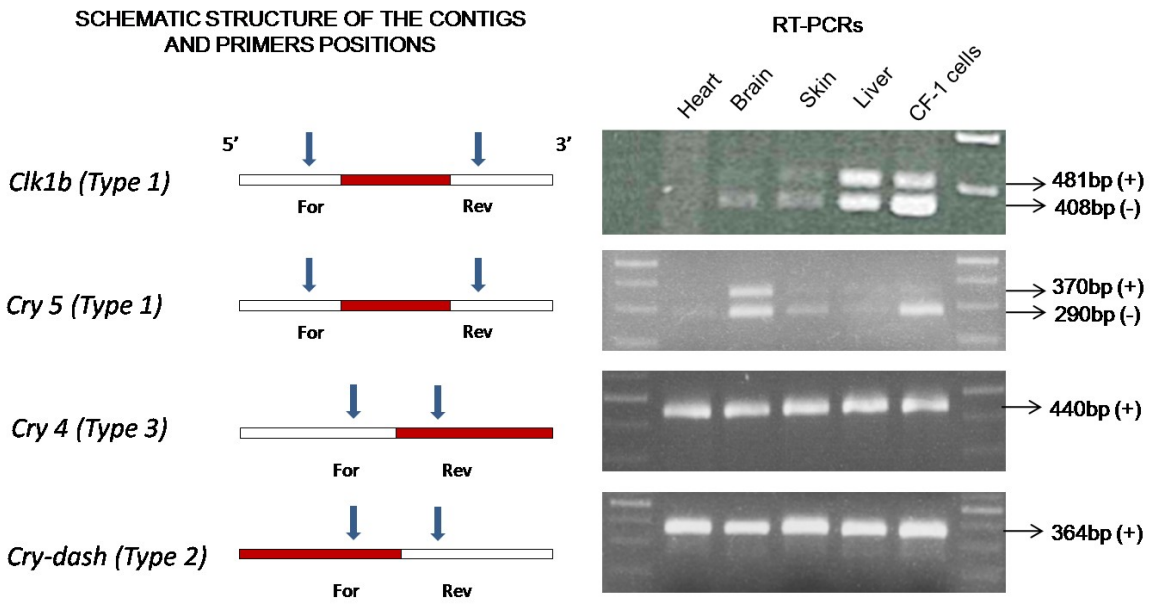

Figure 16. RT-PCR analysis to validate the presence of retained introns in cavefish clock genes. Cavefish RNA was extracted from the CF-1 cell line and from adult heart, brain, skin and liver. Each primer pair was specifically designed in order to detect the presence of the retained introns, on the basis of the structure of each contig predicted from the bioinformatic analysis. For Type 1 contigs, primers flanking the intron were designed on the adjacent exons; for Type 2 contigs, the Forward primer was designed inside the intron sequence and Reverse primer inside the adjacent exon, and *vice versa* for Type 3 contigs. The lengths of the PCR products with (+) and without (-) introns insertions are indicated on the right side of the figure.

The structures and the relative abundance of the various cavefish sequence contigs for each of the clock and visual/non-visual opsin genes, were determined by performing a one-by-one pairwise sequence alignment of each single contig with the homologous sequences from *D. rerio* in GenBank. This "manual" method resulted in a precise description of the intron retention phenomenon, but was also extremely time consuming and applicable only on a small scale. However, in order to fully assess the potential functional consequences of a higher frequency of retained introns in certain genes, it will be crucial to ultimately quantify the level of intron retention at the whole cavefish transcriptome level, through bioinformatics approaches involving the development of specific algorithms. Some preliminary results have been obtained in this regard and will be further described in the Discussion section.

As a next step, we wished to complement our whole transcriptome sequence analysis of *P. andruzzii* by obtaining a first draft whole genome sequence of this species.

3.4 DNAseq analysis in *Phreatichthys andruzzii*

3.4.1 General statistics on *Phreatichthys andruzzii* DNAseq

A total of seven libraries were generated using *P. andruzzii* genomic DNA extracted from the CF-1 cell line. From these libraries, a total number of ~1.5 billion reads have been generated, corresponding to a cumulative size of 142 Gb. The average genome coverage by libraries of DNA fragments, estimated on the basis of the zebrafish genome size (~1,5 Gb, Howe et al., 2013), is ~38 X. The percentage of useful reads from each library has been used for the cavefish *de novo* genome assembly, performed with ALLPATHS-LG software (Gnerre et al., 2010). Zebrafish Refseq transcripts, as well as Trinity transcripts obtained from the cavefish CF-1 cell line and brain RNAseqs, have been mapped on the genomic scaffolds with Blat and then parsed to provide single hits with the highest number of matches. The statistics concerning the genome assembly are shown in Table 3. A major obstacle that limited the generation of our *P. andruzzii* first genome draft data, was the considerable weight that gaps, (present as polyN tracts of undetermined nucleotides), have on the assembly. Partly as a result of this limitation, only a few of the clock genes and light-input pathway elements of interest were completely characterized at the genomic sequence level. Thus, one of the future goals of our research in cavefish is to repeat more DNAseq analysis, attempting to generate longer reads, with a higher sequence accuracy and greater throughput in order to fill as many of the assembly gaps as possible. Nevertheless, the data obtained from this first cavefish whole genome assembly have proved particularly valuable for characterizing the coding sequences of the visual/non-visual opsins which, due to low expression levels, were frequently represented as partial coding sequences in our transcriptome analysis. Furthermore, many visual opsin genes were completely not represented in the transcriptome sequences.

	***P. andruzzii* genome assembly**
Number of contigs	**171,284**
Number of scaffolds	**10,075**
Longest scaffold	**6.3 Mb**
Total assembly size, with gaps	**1 Gb**
N50 contigs (kb)	**5.8**
N50 scaffolds, with gaps (kb)	**755**

Table 3. *P. andruzzii de novo* genome assembly. ALLPATHS-LG software (Broad Institue, MIT) was used to assemble the high-quality reads generated from seven libraries of DNA (two paired-end libraries of DNA fragments and five mate pair libraries). A 1Gb genome assembly draft has been generated, considering also the gaps.

3.4.2 The analysis of the visual/non-visual opsins from the cavefish genome

The assembly of the *P. andruzzii* draft genome allowed us to identify and map on the scaffolds the position of 20 out of 22 of all the different visual and non-visual opsins that have been so far identified in zebrafish (Davies et al., 2011).

The group of visual opsins consists of three different classes having different spectral sensitivities (Bowmaker, 2008): two short wavelength opsins (*Opn1sw1* and *Opn1sw2*) detecting respectively UV and blue light, four medium wavelength opsins (*Opn1mw1, Opn1mw2, Opn1mw3* and *Opn1mw4*) detecting the green region of the visible light and two long wavelength opsins (*Opn1lw1* and *Opn1lw2*) which detect long wavelengths of light including red and yellow. These genes, mainly expressed in the retina and involved in image formation, are well described in zebrafish, while in cavefish they have never been characterized before this DNAseq analysis. The only exceptions that were found expressed in cavefish transcriptome are the short wavelength opsin *Opn1sw2*, obtained in full-length from the brain RNAseq (see Table 2), and the medium wavelength opsin *Opn1mw3*, a short partial fragment again expressed in the brain and showing a retained intron (see Figure 15).

In addition to these visual opsins, we mapped and identified almost all opsins which contribute to peripheral light sensitivity and circadian rhythms, such as melanopsins (*Opn4m1, Opn4m2* and *Opn4m3*), *Xenpous*-like melanopsins (*Opn4x1* and *Opn4x2*), teleost multiple tissue opsin (*TMT-opsin*), rhodopsins and retinal pigment epithelium-

derived rhodopsin (*Rho* and *Rrh*), exorhodopsin (*Exorh*), encephalopsin (*Opn3*), neuropsin (*Opn5*), vertebrate ancient long opsin b (*Valopb*). (Bellingham et al., 2002; Davies et al., 2011; Mano et al., 1999; Moutsaki et al., 2003; Peirson et al., 2009; Pierce et al.,2008). Only two of the complete set of 22 opsins described in zebrafish were not detected. Specifically, the vertebrate ancient long opsin a (*Valopa*) and the retinal G protein coupled receptor a (*Rgra*), possibly as a result of gaps in the genome. However, the full-length coding sequence of these last two genes has been obtained from our RNAseq analysis, revealing that these genes do contribute functional mRNA transcripts (see Table 2).

The many tracts of undetermined sequences in the cavefish genome permitted a fragmentary reconstruction of many of the genes described above. Nevertheless, the mapping of these visual and non-visual opsins on the genomic scaffolds as well as transcriptomic sequences helped us to obtain at least partial exonic sequences for each of these genes. Our next goal was to use these exonic sequence fragments as the basis for designing specific primers for qPCR, in order to measure the levels of mRNA expression of these genes in cavefish in comparison with zebrafish as a control. We attempted to design each pair of primers in sequence regions well conserved between cavefish and zebrafish, with the aim of being able to use the same primers to amplify opsin cDNAs from both species and thereby determine the relative levels of expression. To do this, a pairwise sequence alignment of each the visual and non-visual opsins sequences mapped on the cavefish genome was performed against the corresponding zebrafish homologous coding sequences obtained from GenBank. The short exonic sequence fragments available for the target genes obtained from the cavefish genome, plus the variability of the nucleotide sequences between the two species, in many cases posed a significant technical challenge to design suitable primer sequences. Therefore we focused on a sub-set of genes where effective primer design was possible. Using selected opsin gene–specific primers, we were able to test by qPCR the expression levels of two long wavelengths sensitive (*Opn1lw1* and *Opn1lw2*), one short wavelength sensitive (*Opn1sw2*) visual opsin, plus melanopsin (*Opn4m3*), teleost multiple tissue opsin (*TMT-opsin*), encephalopsin (*Opn3*) and neuropsin (*Opn5*). The cDNA obtained from cavefish and zebrafish brain was used in this analysis, since has been shown that these photoreceptors are typically expressed in the pineal gland and in the deep brain structures in zebrafish (Davies at al., 2011). The results are summarized in Figure 17. Regarding the visual opsins group (Figure 17A), the two long wavelengths sensitive opsins *Opn1lw1* and *Opn1lw2* show extremely low expression levels, particularly in cavefish brain, consistent with the lack of these sequences in our RNAseq analysis or former manual cloning attempts.. Surprisingly, the short wavelength

sensitive opsin *Opn1sw2* showed much higher expression levels in cavefish than in zebrafish brain. Expression of the non-visual opsins (Figure 17B) showed a clear trend in which each of the genes analyzed has higher expression levels in the zebrafish than in cavefish brain. As can be seen from the two graphs, the relative mRNA levels of the non-visual opsins is ~100 times higher in the zebrafish and cavefish brain samples, compared to the relative mRNA levels of the visual opsins in the same tissues. The reason could be the fact that the visual opsins are mainly expressed in the retina rather than the brain, in contrast with the non-visual opsins group (Davies at al., 2011).

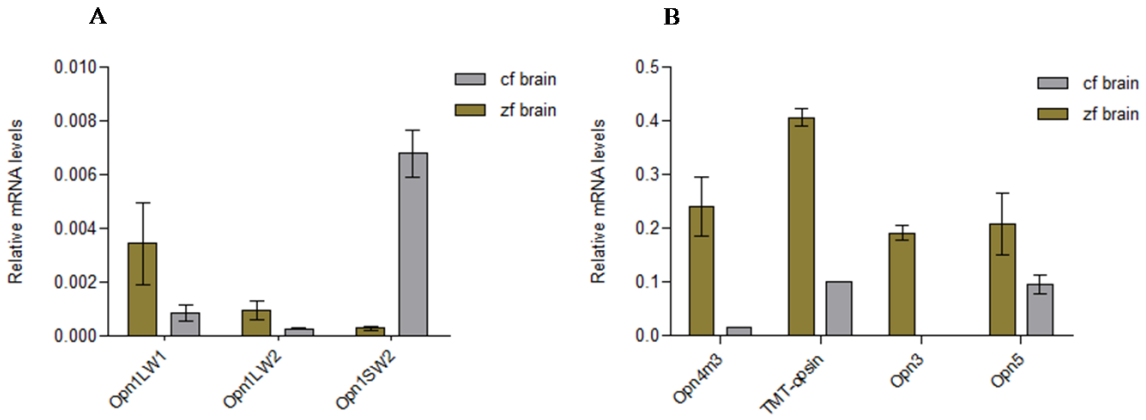

Figure 17. Visual and non-visual opsin expression levels in cavefish and zebrafish brain. qPCR analysis of the visual opsins (A) shows very low expression levels of the two long wavelength sensitive opsins *Opn1lw1* and *Opn1lw2* and a higher expression level of the short wavelength sensitive opsin *Opn1sw2* in the brain of *P.andruzzii*. For the non-visual opsins (B), the levels of expression are clearly higher in zebrafish brain, compared with the cavefish. Mean and SEM plotted on the graph. The graphs were plotted with GraphPad Prism 5.0 (GraphPad Software Inc.).

The results obtained by the qPCR analysis, specifically about the expression levels of the three visual opsins *Opnlw1*, *Opnlw2* and *Opnsw2*, are important as they may help to explain some outcomes from the behavioral experiments that we performed in the University of Ferrara on adult cavefish exposed to light-dark cycles (LD) under different wavelengths of light. In these behavioral experiments, it has been measured the locomotor activity of adult fish exposed to a 12:12 hours LD cycle, under different wavelengths including red and blue light. To measure the movement of the fish, infrared photocells were set on the aquarium walls and connected to a computer. The activity is illustrated by a so-called actogram, where movements of the fish swimming through the infrared beam are

recorded and plotted. X^2 periodograms (Sokolove-Bushell test) were also performed to test whether the temporal series of locomotor activity data show a periodic component or not. The results of the analysis regarding the exposure of the fish under red and blue light are shown in Figure 18.

On the left (Figure 18A) is shown the actogram and the periodograms related to the blue light exposure experiment. After a period of constant light (LL), which shows a total arrhythmic locomotor activity pattern, the fish were exposed to three periods of LD cycles under blue light: the LD1 cycle, then the LD2 cycle, where the lights were switched on with a delay of eight hours (shifted forward), and the LD3 cycle, in which the light-dark phase were brought back of eight hours as in LD1. A robust rhythmicity has been detected in these three LD phases, showing a strong photophobic response of the fish during the 12 hours of light with a reduction of the locomotor activity, and a subsequent increasing of the activity in the next 12 hours of dark. Also the periodograms, related to the LD2 cycle as example, show a strong periodicity of 24 hours in the locomotor activity of the cavefish. At the end, a final period of constant darkness (DD) shows again a total arrhythmic locomotor activity pattern as in the LL phase. Lastly, on the right (Figure 18B) is shown the actogram and the periodograms related to the red light exposure experiment. As can be seen, in this experiment we detected arrhythmic locomotor activity in all the light cycles (LL, LD1, LD2, LD3 and DD) and the periodograms show no periodic pattern of activity, in contrasts to the robust period observed in the blue light exposure experiment.

The results of these behavioral experiments seem to be in agreement with the qPCR tests performed on the visual opsins *Opn1lw1*, *Opn1lw2* and *Opn1sw2*. The short wavelength sensitive opsin *Opn1sw2*, which shows very high expression levels in cavefish brain, could represent one of the candidates that might mediate the detection of the blue light, giving as a response the strong periodic rhythm of locomotor activity observed in the fish exposed under these wavelength cycles. Conversely, the extremely low levels of expression of the two long wavelengths sensitive opsins *Opn1lw1* and *Opn1lw2*, which have a sensitivity in the red region of the electromagnetic spectrum, might explain the lack of rhythmicity in locomotor activity of the fish exposed under red light cycles, proving that the cavefish may not detect this range of long wavelengths.

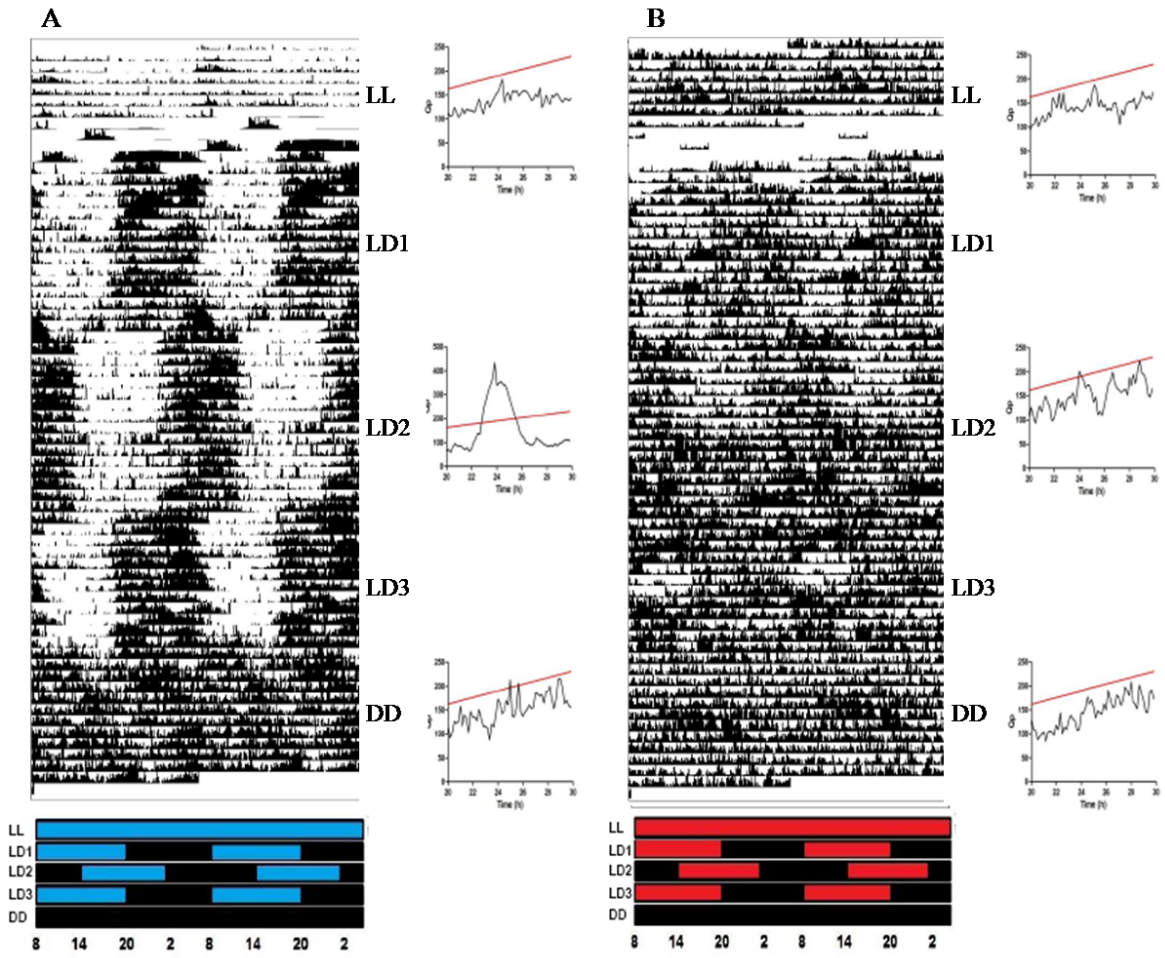

Figure 18. Locomotor activity experiments on adult *P. andruzzii* exposed under blue and red light cycles. Arrhythmic locomotor activity was detected in all the cycles of red light, proving that this wavelength does not have an influence on the fish. Conversely, a strong rhythmic pattern of activity has been shown in the fish exposed to 12:12h LD cycle under blue light. The periodogram in LD2, as well as LD1 and LD3 shows a clear peak representing a 24h period. Actograms and periodograms were plotted with the chronobiology software El Temps (version 1.228).

IV. Discussion

4.1 Cavefish *Period2*: a candidate underlying the aberrant circadian clock properties

One major goal of our study was the characterization of the cavefish *P. andruzzii* circadian clock genes in order to find mutations that could explain the abnormal clock phenotype of this species. Cloning the cavefish clock elements by homology with their zebrafish counterparts has revealed significant mutations in one key core clock gene, *Period2*. The analysis of *Period2* revealed the insertion of a transposable element in-frame in the coding sequence, located after the CK1ε phosphorylation site which is responsible for regulating PER2 degradation, resulting in the addition of a novel 75 amino acid sequence to the PER2 protein. The presence of the transposon, first detected in the cavefish cell line, has also been confirmed *in vivo* in different tissues of the animal. Furthermore, a premature stop codon mutation was found downstream of the transposon in the C-terminal region of the coding sequence, which gives as a result the deletion of the putative binding site for CRY involved in the translocation of the PER:CRY complex into the nucleus (Tomita et al., 2010). Together, these two mutations could lead to a change in the three-dimensional structure of the PER2 protein, which might disrupt the stability and turnover of this protein as well as its ability to interact with other components of the oscillator. Interestingly, the earliest studies aiming to identify clock genes in *Drosophila* revealed that mutations leading to shortening or lengthening of the free-running period length were all attributable to structural changes or loss of function of the *Period* gene (Konopka & Benzer, 1971; Reddy et al., 1984; Zehring et al., 1984).

Our analysis of cavefish *Per2* was achieved by comparing the sequence of the gene's full-length Open Reading Frame (ORF) with the complete genomic structure, obtained by manual cloning of the introns as well as whole genome sequence data analysis. The comparison of the expression levels of *Per2* and *Cry1a* in cavefish and zebrafish, gave us important hints as to the functional contribution of these two genes. *Cry1a* levels of expression are significantly higher than *Per2* in every tissue analyzed in both species, suggesting a possible major role of this gene in the core clock mechanism. Interestingly, in contrast to *Per2*, the cavefish *Cry1a* gene does not show any mutations and the alignment

of its full ORF with its zebrafish homolog reveals high conservation along its entire length. One important goal for the future is to compare in detail the sub-cellular localization of the cavefish and zebrafish PER2 proteins by immunostaining experiments, with the aim of better understanding how cavefish PER2 is able to function as an inhibitor of the CLK:BMAL complex and how the different mutations encountered in the cavefish gene can affect its normal regulatory function in the clock mechanism.

4.2 RNAseq analysis and intron retention in *Phreatichthys andruzzii* mRNA

The analysis of cavefish *Period2* and its mutations prompted us to explore whether other mutations might affect the clock gene coding sequences, as well as the genes involved in photoreception and the light input pathway. In order to address this question we chose to adopt a simultaneous screening approach that could help us to obtain a complete map of the elements involved in the cellular regulation of the circadian clock and their possible mutations. Thanks to the recent and rapid development of Next Generation Sequencing (NGS) tools, we decided to apply a global strategy in *P. andruzzii* by transcriptome analysis (RNAseq). By generating two *de novo* transcriptome assemblies from the CF-1 cell line and brain RNA, we provided for the first time a statistical and functional description of *P. andruzzii* transcriptome, as well as a detailed sequence analysis of clock genes and genes involved in photoreception and light input pathway.

Indications about the quality of our assemblies can be deduced by comparing some basic statstics from our study on *P. andruzzii* with two different analysis performed on the recently published *Astyanax mexicanus* transcriptome, even though different sequencing techniques and *de novo* assembly approaches have been adopted in these two cases. In the first study (Hinaux et al., 2013), *de novo* Sanger sequencing of embryonic and larval transcriptomes of *A. mexicanus* cave and surface-dwelling fish has been presented. Clones of libraries have been sequenced and assembled in a total of 44,145 contigs with a mean length of 985 bp, while using a Blastx search against the zebrafish proteome, these sequences corresponded to 11,197 different proteins. The second study (Gross et al., 2013) concerns an integrated *de novo* transcriptome assembly of *A. mexicanus* from adult cave and surface-dwelling fish using Roche/454 sequencing technology, generating a total of 22,596 assembled contigs, with a mean length of 1,406 bp and a N50 of 1,781 bp. More

technically similar to our analysis performed on *P. andruzzii* is the characterization of the *Poecilia mexicana* transcriptome from adult fish gill, by Illumina sequencing technology and *de novo* assembly with Trinity software (Kelley et al., 2012). *P. mexicana*, the Atlantic molly, is an emerging model system useful to study adaptation to extreme environments and ecological speciation, as the species has colonized both hydrogen sulfide rich and cave habitats in southern Mexico, exhibiting a wide range of extreme physical properties similar to *P. andruzzii* and *A. mexicanus*. In this study, a total number of 80,111 contigs have been obtained, with a mean length of 932 bp and 17,814 proteins have been annotated against SwissProt database. As presented in the Result section, our first RNAseq analysis by Illumina sequencing technology and *de novo* Trinity assembly of *P. andruzzii* transcriptome, resulted in a total number of 272,717 contigs assembled for the CF-1 cell line, with a mean length of 1,065 bp and N50 of 2,187 bp, while a total number of 7,167 proteins have been annotated. Regarding the brain analysis, a total number of 380,678 contigs have been assembled, with a mean length of 800 bp and N50 of 1,635 bp, while from the KOG analysis, a total number of 10,634 proteins have been annotated (Table 4). The comparison between the statistics obtained from our *P. andruzzii* transcriptome analysis and the other cases above described, point to a relatively high quality for our first *de novo* assemblies, which resulted in reliably comparable datasets with other recently generated fish transcriptomes, even though using different sequencing and assembly methods.

	***P. andruzzii* CF-1 cell line**	***P. andruzzii* Brain**	***A. mexicanus* (Hinaux et al., 2013)**	***A. mexicanus* (Gross et al., 2013)**	***P. mexicana* (Kelley et al., 2012)**
Number of contigs	**272,717**	**380,678**	**44,145**	**22,596**	**80,111**
Average contigs length (bp)	**1,065**	**800**	**985**	**1,406**	**932**
N50 contigs (kb)	**2,187**	**1,635**	**not indicated**	**1,781**	**not indicated**
Number of annotated proteins	**7,167**	**10,634**	**11,197**	**not indicated**	**17,814**

Table 4. RNAseq statistics. Summary statistics from different transcriptome analysis in *P. andruzzii*, *A. mexicanus* and *P. mexicana*.

During the characterization of the clock genes and visual/non-visual opsins obtained by RNAseq from the CF-1 cell line and cavefish brain we were able to reveal a major finding, namely the high rate of unspliced introns retained inside certain categories of genes, which in many cases introduce premature stop codons leading to truncated proteins. Deep Sequencing tools developed in recent years has revealed many fascinating but poorly understood phenomena, including intron retention in processed eukaryotic transcripts. The intron retention in mRNAs has been first described in the context of the Nonsense-Mediated mRNA Decay (NMD) mechanism, a quality-control mechanism that selectively degrades mRNAs harboring premature termination (nonsense) codons (Chang et al., 2007). Very recently, a detailed study has revealed the widespread presence of retained introns in mammalian transcriptomes although at a lower frequency than that we have described here in *P. andruzzii*. It has been speculated that the regulation of the level of intron retention represents a potential mechanism for functionally "fine tuning" transcriptomes (Braunschweig et al., 2014). This study revealed that the mechanism of intron retention (IR) can be classified as a subset of the canonical, well-studied alternative splicing (AS) mechanisms, which are responsible for expanding the regulatory and functional capacity of eukaryotic genomes. This IR mechanism acts widely in order to reduce the levels of transcripts that are less or not required for the physiology of a cell or tissue type in which they are detected, through NMD as well as nuclear sequestration and turnover of IR transcripts (Braunschweig et al., 2014). Moreover, different classes of retained introns have been discovered in this study, as we could also detect from our results on *P. andruzzii* RNAseq analysis. It is tempting to speculate that the intron retention phenomenon we have described in *P. andruzzii* mRNA may explain the aberrant clock phenotype of this species as well as potentially contributing to other extreme troglomorphic phenotypes.

Our "manual" method of analysing intron splicing of individual genes in the transcriptome was very effective in revealing the intron retention phenomenon, but was also extremely time consuming and applicable only to a small sub-set of genes. For this reason, the next goal of our research is to try to extend the screening for contigs with retained introns at the whole transcriptome level, through the use of bioinformatic algorithms that can automatically screen transcript sequences for unspliced introns. In collaboration with the Shanghai Institute of Computational Biology we have now developed a first basic algorithm in order to automatically scan the whole transcriptome for retained introns. The algorithm is based on the alignment of all cavefish transcripts with the zebrafish reference genome as well as our first *P. andruzzii* draft genome, in order to compare the intron-exon junctions positions and identify unspliced introns. This method has been applied to the CF-

1 cell line transcriptome, while its application on the brain transcriptome is currently ongoing. As a first result, a total of around 4,000 sequences have been marked as possible candidates for intron retention, thus revealing a surprisingly large number of affected genes as well as a specificity for certain classes of genes such as the clock genes group or the elements involved in the intron splicing machinery and DNA-excision repair mechanism.
The next task in our cavefish sequence analysis is to improve this algorithm in order to minimize the error values and facilitate its application to a large number of RNAseq datasets from different *P. andruzzii* tissues. In this regard, it is interesting to note that the the *de novo* Trinity assembly approach that lead us to the first identification of intron retaining contigs, is useful in a qualitative way but not for the precise quantification of intron retention, given the fact that the contigs are generated by consolidating a large number of raw sequencing reads. For this reason, the alignment of transcripts as raw reads onto the zebrafish/cavefish genomes and the application of a more refined algorithm to analyse this larger dataset will be required. The confirmation of these findings will be of fundamental importance in order to proceed further in detailed functional experiments. We are currently embarking on studies of the functional consequence of certain genes being strongly affected by intron retention, for example *Cry5* and *Cry-dash* which are responsible for the repair of UV-damaged DNA, the candidate photoreceptor *Cry4*, circadian clock elements and spliceosome encoding genes. Our challenge is to link this phenomenon with specific phenotypic changes in response to a set of extreme environmental conditions, such as altered metabolic processes, disruption in light-detection pathways and DNA repair mechanisms or loss of eyes and pigmentation. By this approach, we will ultimately be able to assess whether alterations in the efficiency of the splicing machinery could serve as an evolutionary mechanism for adapting to extreme environmental conditions.

4.3 Genome sequencing and visual/non-visual opsins analysis in *Phreatichthys andruzzii*

In parallel with the RNAseq analysis performed on the CF-1 cell line and cavefish brain, we also obtained a first draft genome of *P. andruzzii* by DNAseq and genomic *de novo* assembly. By this analysis, a draft genome of this species has been obtained and assembled in a total size of around 1Gb, with gaps. In order to have an indication of the quality of our

P. andruzzii draft genome assembly, we tried to compare its parameters with two new draft genomes recently sequenced by Illumina technology and assembled with ALLPATHS-LG software as in our case, that is *Latimeria chalumnae* and *Astyanax mexicanus* genomes.
L. chalumnae (the coelacanth) is the only living member of an ancient group of lobe-finned fishes that was known previously only from fossils and believed to have been extinct since the Late Cretaceous period (Amemiya et al., 2013). Its genome draft has been assembled to a size of 2.86 Gb, considering also gaps, and shows a contig N50 size of 12.7 kb and a scaffold N50 size of 924 kb (Amemiya et al., 2013). In *A. mexicanus* the genome draft has been assembled to a size of 964 Mb, with gaps, and shows a contig N50 size of 14.7 kb and a scaffold N50 size of 1,775 kb (McGaugh et al., 2014). Thus, our *P. andruzzii* draft assembly (1 Gb) is more similar in length to *A. mexicanus* (964 Mb) rather than *L. chalumnae* (2.86 Gb), while regarding the N50 size of the contigs and scaffolds, these parameters result lower in *P. andruzzii* compared to the other assemblies (Table 5).

	P. andruzzii	***A. mexicanus***	***L. chalumnae***
Total assembly size, with gaps (Gb)	**1**	**0.964**	**2.86**
N50 contigs (kb)	**5.8**	**14.7**	**12.7**
N50 scaffolds, with gaps (kb)	**755**	**1,775**	**924**

Table 5. DNAseq statistics. Summary statistics from genome sequencing analysis in *P. andruzzii*, *A. mexicanus* and *L. chalumnae*.

Gaps in the assembly induced inevitable problems in the handling of our *P. andruzzii* first genome draft data, and so resulted in a relatively fragmentary genome assembly. One of the future goals of our study in cavefish is to repeat more DNAseq analysis, trying to generate longer reads, with a higher sequence accuracy and greater throughput in order to fill he gaps as much as possible. A more accurate genome assembly will be a fundamental requirement for the analysis of intron retention in *P. andruzzii*. Indeed, by reducing as much as possible gaps in the assembly, this would allow us to directly align the transcriptomic sequence data on the exact genomic sequence of the species, avoiding the use of zebrafish genome as a reference, and thus reconstructing precisely the intron-exon junctions positions and detecting accurately the unspliced introns in the cavefish transcripts. Nevertheless, the data obtained from the cavefish DNAseq analysis were extremely informative for our detailed study of the visual/non-visual opsins in *P. andruzzii*. By this analysis, we identified and mapped on the cavefish genome the position of 20 out

of 22 of all the visual and non-visual opsins that have been so far identified in zebrafish (Davies et al., 2011).

Interestingly, in all the partial coding sequences identified in the cavefish genome, no point mutations have been encountered, with the exception of the previously identified mutated opsins *TMT-opsin* and *Opn4m2*. Based on these results, we can speculate that the loss of photoreceptor function in cavefish may result either from mutations affecting the coding regions of opsins (as in *TMT-opsin* and *Opn4m2*) or, alternatively, from mutations affecting the levels of opsin gene transcription. This may be the case for example in the two long wavelength sensitive opsins *Opn1lw1* and *Opn1lw2*; the extremely low expression levels measured by qPCR could explain the lack of response in cavefish to the range of long wavelengths (red-yellow light) as well as explaining why these genes were not represented in our RT-PCR or RNAseq-based analysis. Conversely, the short wavelength sensitive opsin *Opn1sw2* expressed at relatively high levels in the cavefish brain might mediate the strong behavioral response of the fish under blue light. More detailed DNAseq analysis will be required in order to obtain longer or more complete opsin coding sequences for further studies, such as completing our search for point mutations and generating more stringent primers for qPCR analysis. It is tempting to speculate that plasticity in opsin gene expression control mechanisms may enable cavefish to adapt to changes in their photic environment during evolution. A promoter region analysis of the opsin genes and a comparison with their zebrafish counterparts will be also important to test whether key promoter enhancer motifs are absent or mutated, and so may explain the species–specific differences in expression levels.

4.4 *Phreatichthys andruzzii* and its aberrant clock phenotype

The 24 hour day-night cycle strongly influences life on Earth and is predicted to have lead to the evolution of intrinsic circadian oscillators. These allowed plants and animals to predict and adapt to the regular changes in their environment. Likewise, the evolution of *Phreatichthys andruzzii* has been strongly influenced by its isolation from the day-night cycle during the last 1.4-2.6 million years in deep cave systems under the Somalian desert. The most extreme adaptions to life in this dark environment are the loss of eyes, scales, complete depigmentation and reduced metabolism. Furthermore, an aberrant circadian clock has also been discovered in this species. Data from our group has revealed that, while in most organisms the clock is entrained by light-dark cycles, the clock of *P.*

andruzzii is no longer entrainable by this signal. In contrast, upon exposure to alternative zeitgebers such as dexamethasone in cell-cultured lines, the cavefish clock cycles with an unexpected long infradian period of longer than 40 hours and shows reduced temperature compensation (Cavallari et al., 2011). This phenomenon could either reflect the progressive loss of a mechanism that confers no selective advantages for animals that live under constant conditions of darkness and temperature or, as an alternative explanation, the loss of a normal circadian clock phenotype may conversely represent a mechanism under positive selection. This hypothesis would predict the existence of an infradian rhythm of environmental signals in the cave system, therefore abnormal properties such as an infradian free-running period might confer some selective advantages on the fish in this environment. A recent study on the circadian rhythms in *Astyanax mexicanus* revealed some differences between the clock mechanism of this Mexican species and the Somalian *P. andruzzii*. Upon exposure to LD cycles, *A. mexicanus* adult surface fish show at the molecular level robust circadian rhythms in *Per1* expression, which are also retained in cave populations but with substantial alterations such as lower amplitude and shifted peak of expression, showing that the cavefish populations still retain the ability to detect light and generate molecular circadian oscillations (Beale et al., 2013). Furthermore, at the behavioral level, Mexican cavefish locomotor activity in LD is weakly rhythmic with a general increase in the day, while clock-controlled locomotor rhythms are absent in cavefish in DD following entrainment by exposure to a LD cycle (Beale et al., 2013). These findings suggest that the circadian phenotype observed in *Astyanax* cavefish is much less drastic than in the Somalian *P. andruzzii*, which has completely lost the ability to entrain to a LD cycle probably due to a much longer evolutionary time in complete darkness.

Another important finding in *P. andruzzii* is that, upon exposure to a regular 24 hour feeding-timed cycle, the cavefish exhibit a normal circadian activity rhythm, demonstrating thus the presence in *P. andruzzii* of a circadian clock mechanism entrainable by food's administration and not by LD cycles. This result implies that food availability in the subterranean environment of this cavefish might indeed be present with a periodicity of 24 hours, therefore a clock responding and anticipating this signal may confer a survival advantage. However, the difficulty to study the ecology of this subterranean environment in the Somalian desert makes it difficult to test this hypothesis. Feeding entrainment of the circadian timing system is documented in several studies, which highlights the importance of food as a zeitgeber in fish as well as the existence of a food-entrainable oscillator (FEO), which is genetically distinct from the normal light entrainable oscillator (LEO)

(Lopez- Olmeda & Sanchez-Vazquez, 2010; Lopez-Olmeda et al., 2010). Comparative studies between zebrafish, in which both normal light and food entrainable oscillators are present, and *P. andruzzii*, which shows only a normal food entrainable oscillator, could help us to provide important insights into the basis of the feeding entrainment mechanism in vertebrates. Up to now, the molecular mechanism and the anatomical location of the FEO remains unclear (Escobar et al., 2009; Lopez-Olmeda et al., 2010; Stephan, 2002; Storch & Weitz, 2009).

In our studies on the cavefish *P. andruzzii* we have documented the existence, in the same species, of an infradian clock measured at the molecular level in a cell culture model and a circadian, food-regulated clock revealed at the behavioral level. It is classically assumed that the basic circadian clock parameters measured in a primary cell culture accurately reflect the clock properties manifest at the whole animal level. In the case of *P. andruzzii*, our results display an inherent complexity within the circadian clock system, which can accommodate the coexistence of clocks with different period lengths that respond to different signals, predicting thus a possible hierarchy of distinct clocks that may differ depending on the zeitgeber. From all these findings, in addition to displaying a unique and interesting collection of adaptations to its extreme environment, *P. andruzzii* proved to be a powerful model for studying the evolution of the circadian systems in extreme environments as well as the evolution of the many pathways and processes in response to light.

V. References

Albrecht, U., Bordon, A., Schmutz, I., & Ripperger, J. (2007). The multiple facets of *Per2*. Cold Spring Harb Symp Quant Biol, 72, 95-104.

Albrecht, U., Zheng, B., Larkin, D., Sun, Z. S., & Lee, C. C. (2001). M*Per1* and m*Per2* are essential for normal resetting of the circadian clock. J Biol Rhythms, 16(2), 100-104.

Albus, H., Vansteensel, M. J., Michel, S., Block, G. D., & Meijer, J. H. (2005). A GABAergic mechanism is necessary for coupling dissociable ventral and dorsal regional oscillators within the circadian clock. Curr Biol, 15(10), 886-893.

Amemiya, C. T., Alfoldi, J., Lee, A. P., Fan, S., Philippe, H., MacCallum, I., et al. (2013). The African coelacanth genome provides insights into tetrapod evolution. Nature, 496, 311-316.

Antle, M. C., & Silver, R. (2005). Orchestrating time: arrangements of the brain circadian clock. Trends Neurosci, 28(3), 145-151.

Appelbaum, L., Toyama, R., Dawid, I. B., Klein, D. C., Baler, R., & Gothilf, Y. (2004). Zebrafish serotonin-N-acetyltransferase-2 gene regulation: pineal-restrictive downstream module contains a functional E-box and three photoreceptor conserved elements. Mol Endocrinol, 18(5), 1210-1221.

Appelbaum, L., Vallone, D., Anzulovich, A., Ziv, L., Tom, M., Foulkes, N. S., & Gothilf, Y. (2006). Zebrafish arylalkylamine-N-acetyltransferase genes -targets for regulation of the circadian clock. J Mol Endocrinol, 36(2), 337-347.

Aspiras, A. C., Prasad, R., Fong, D. W., Carlini, D. B., & Angelini, D. R. (2012). Parallel reduction in expression of the eye development gene hedgehog in separately derived cave populations of the amphipod *Gammarus minus*. J Evol Biol, 25(5), 995-1001.

Aton, S. J., Colwell, C. S., Harmar, A. J., Waschek, J., & Herzog, E. D. (2005). Vasoactive intestinal polypeptide mediates circadian rhythmicity and synchrony in mammalian clock neurons. Nat Neurosci, 8(4), 476-483.

Bae, K., Jin, X., Maywood, E. S., Hastings, M. H., Reppert, S. M., & Weaver, D. R. (2001). Differential functions of m*Per1*, m*Per2*, and m*Per3* in the SCN circadian clock. Neuron, 30(2), 525-536.

Balsalobre, A., Damiola, F., & Schibler, U. (1998). A serum shock induces circadian gene expression in mammalian tissue culture cells. Cell, 93(6), 929-937.

Beale, A., Guibal, C., Tamai, K. T., Klotz, L., et al. (2013) Circadian rhythms in Mexican blind cavefish *Astyanax mexicanus* in the lab and in the field. Nat. Commun. 4:2769 doi: 10.1038/ncomms3769.

Bellingham, J., Whitmore, D., Philp, A. R., Wells, D. J., & Foster, R. G. (2002). Zebrafish melanopsin: isolation, tissue localisation and phylogenetic position. Brain Res Mol Brain Res, 107(2), 128-136.

Ben-Moshe, Z., Vatine, G., Alon, S., Tovin, A., Mracek, P., Foulkes, N. S., & Gothilf, Y. (2010). Multiple PAR and E4BP4 bZIP transcription factors in zebrafish: diverse spatial and temporal expression patterns. Chronobiol Int, 27(8), 1509-1531.

Berson, D. M., Dunn, F. A., & Takao, M. (2002). Phototransduction by retinal ganglion cells that set the circadian clock. Science, 295(5557), 1070-1073.

Berti, R., Durand, J. P., Becchi, S., Brizzi, R., Keller, N., & Ruffat, G. (2001). Eye degeneration in the blind cave-dwelling fish *Phreatichthys andruzzii*. Canadian Journal of Zoology, 79(7), pp. 1278-1285(1278).

Bjarnason, G. A., & Jordan, R. (2000). Circadian variation of cell proliferation and cell cycle protein expression in man: clinical implications. Prog Cell Cycle Res, 4, 193-206.

Borowsky, R. (2008). Restoring sight in blind cavefish. Curr Biol, 18(1), R23-24.

Borrelli, E., Nestler, E. J., Allis, C. D., & Sassone-Corsi, P. (2008). Decoding the epigenetic language of neuronal plasticity. Neuron, 60(6), 961-974.

Bowmaker, J. K. (2008). Evolution of vertebrate visual pigments. Vision Res, 48(20), 2022-2041.

Braunschweig, U., Barbosa-Morais, N. L., Pan, Q., Nachman, E. N., Alipanahi, B., Gonatopoulos-Pournatzis, T., Frey, B., Irimia, M., Blencowe, B. J. (2014). Widespread intron retention in mammals functionally tunes transcriptomes. Genome Res, 24, 1774–1786. doi: 10.1101/gr.177790.114.

Brittijn, S. A., Duivesteijn, S. J., Belmamoune, M., Bertens, L. F., Bitter, W., de Bruijn, J. D., Champagne, D. L., Cuppen, E., Flik, G., Vandenbroucke-Grauls, C. M., Janssen, R. A., de Jong, I. M., de Kloet, E. R., Kros, A., Meijer, A. H., Metz, J. R., van der Sar, A. M., Schaaf, M. J., Schulte-Merker, S., Spaink, H. P., Tak, P. P., Verbeek, F. J., Vervoordeldonk, M. J., Vonk, F. J., Witte, F., Yuan, H., & Richardson, M. K. (2009). Zebrafish development and regeneration: new tools for biomedical research. Int J Dev Biol, 53(5-6), 835-850.

Brown, S. A., Zumbrunn, G., Fleury-Olela, F., Preitner, N., & Schibler, U. (2002). Rhythms of mammalian body temperature can sustain peripheral circadian clocks. Curr Biol, 12(18), 1574-1583.

Buhr, E. D., Yoo, S. H., & Takahashi, J. S. (2010). Temperature as a universal resetting cue for mammalian circadian oscillators. Science, 330(6002), 379-385.

Bunger, M. K., Wilsbacher, L. D., Moran, S. M., Clendenin, C., Radcliffe, L. A., Hogenesch, J. B., Simon, M. C., Takahashi, J. S., & Bradfield, C. A. (2000). *Mop3* is an essential component of the master circadian pacemaker in mammals. Cell, 103(7), 1009-1017.

Bünning, E. (1935). Zur Kenntnis der erblichen Tagesperiodizität bei den Primärblättern von Phaseolus multifloras. Jahrbuch der Wissenschaftlichen Botanik, 81, 411-418.

Cahill, G. M. (1996). Circadian regulation of melatonin production in cultured zebrafish pineal and retina. Brain Res, 708(1-2), 177-181.

Cahill, G. M. (2002). Clock mechanisms in zebrafish. Cell Tissue Res, 309(1), 27-34.

Carr, A. J., & Whitmore, D. (2005). Imaging of single light-responsive clock cells reveals fluctuating free-running periods. Nat Cell Biol, 7(3), 319-321.

Cavallari, N., Frigato, E., Vallone, D., Frohlich, N., Lopez-Olmeda, J. F., Foa, A., Berti, R., Sanchez-Vazquez, F. J., Bertolucci, C., & Foulkes, N. S. (2011). A blind circadian clock in cavefish reveals that opsins mediate peripheral clock photoreception. PLoS Biol, 9(9), e1001142.

Cermakian, N., Whitmore, D., Foulkes, N. S., & Sassone-Corsi, P. (2000). Asynchronous oscillations of two zebrafish CLOCK partners reveal differential clock control and function. Proc Natl Acad Sci U S A, 97(8), 4339-4344.

Chang, J. P., & Peter, R. E. (1983). Effects of pimozide and des Gly10,[D-Ala6]luteinizing hormone-releasing hormone ethylamide on serum gonadotropin concentrations, germinal vesicle migration, and ovulation in female goldfish, *Carassius auratus*. Gen Comp Endocrinol, 52(1), 30-37.

Chang, Y. F., Imam, J. S., Wilkinson, M. F. (2007). The nonsense-mediated decay RNA surveillance pathway. Annu Rev Biochem advance online publication, doi: 10.1146/annurev.biochem.76.050106.093909.

Chatterjee, S., & Lufkin, T. (2012). Regulatory genomics: Insights from the zebrafish. Curr Top Genet, 5, 1-10.

Cho, H., Zhao, X., Hatori, M., Yu, R. T., Barish, G. D., Lam, M. T., Chong, L. W., DiTacchio, L., Atkins, A. R., Glass, C. K., Liddle, C., Auwerx, J., Downes, M., Panda, S., & Evans, R. M. (2012). Regulation of circadian behaviour and metabolism by REV-ERB-alpha and REV-ERB-beta. Nature, 485(7396), 123-127.

Colli, L., Paglianti, A., Berti, R., Gandolfi, G., & Tagliavini, J. (2009). Molecular phylogeny of the blind cavefish *Phreatichthys andruzzii* and *Garra barreimiae* within the family Cyprinida. Environ Biol Fish, 84(95-107).

Colwell, C. S. (2000). Rhythmic coupling among cells in the suprachiasmatic nucleus. J Neurobiol, 43(4), 379-388.

Cowell, I. G. (2002). E4BP4/NFIL3, a PAR-related bZIP factor with many roles. Bioessays, 24(11), 1023-1029.

Crews, S. T., & Fan, C. M. (1999). Remembrance of things PAS: regulation of development by bHLH-PAS proteins. Curr Opin Genet Dev, 9(5), 580-587.

Crosio, C., Cermakian, N., Allis, C. D., & Sassone-Corsi, P. (2000). Light induces chromatin modification in cells of the mammalian circadian clock. Nat Neurosci, 3(12), 1241-1247.

Culver, D. C. (1982). Cave Life: Evolution and Ecology. Cambridge, MA: Harvard University Press, 189 pp.

Culver, P. D., Humphreys, W. F. , & Langecker, T. G. (2000). Ecosystem of the wolrd 30: The effect of continuous darkness on cave ecology. Ecosystem of the world, pp.135-157.

Dahm, R., & Geisler, R. (2006). Learning from small fry: the zebrafish as a genetic model organism for aquaculture fish species. Mar Biotechnol (NY), 8(4), 329-345.

Damiola, F., Le Minh, N., Preitner, N., Kornmann, B., Fleury-Olela, F., & Schibler, U. (2000). Restricted feeding uncouples circadian oscillators in peripheral tissues from the central pacemaker in the suprachiasmatic nucleus. Genes Dev, 14(23), 2950-2961.

Davies, W. I., Zheng, L., Hughes, S., Tamai, T. K., Turton, M., Halford, S., Foster, R. G., Whitmore, D., & Hankins, M. W. (2011). Functional diversity of melanopsins and their global expression in the teleost retina. Cell Mol Life Sci, 68(24), 4115-4132.

Dekens, M. P., Santoriello, C., Vallone, D., Grassi, G., Whitmore, D., & Foulkes, N. S. (2003). Light regulates the cell cycle in zebrafish. Curr Biol, 13(23), 2051-2057.

Delaunay, F., Thisse, C., Thisse, B., & Laudet, V. (2003). Differential regulation of *Period 2* and *Period 3* expression during development of the zebrafish circadian clock. Gene Expr Patterns, 3(3), 319-324.

Dibner, C., Schibler, U., & Albrecht, U. (2010). The mammalian circadian timing system: organization and coordination of central and peripheral clocks. Annu Rev Physiol, 72, 517-549.

Dickmeis, T. (2009). Glucocorticoids and the circadian clock. J Endocrinol, 200(1), 3-22.

Dowling, T. E., Martasian, D. P., & Jeffery, W. R. (2002). Evidence for multiple genetic forms with similar eyeless phenotypes in the blind cavefish, *Astyanax mexicanus*. Mol Biol Evol, 19(4), 446-455.

Duez, H., & Staels, B. (2008). The nuclear receptors Rev-erbs and RORs integrate circadian rhythms and metabolism. Diab Vasc Dis Res, 5(2), 82-88.

Dunlap, J. C. (1999). Molecular bases for circadian clocks. Cell, 96(2), 271-290.

Dunlap, J. C., & Loros, J. J. (2004). The neurospora circadian system. J Biol Rhythms, 19(5), 414-424.

Dziema, H., Oatis, B., Butcher, G. Q., Yates, R., Hoyt, K. R., & Obrietan, K. (2003). The ERK/MAP kinase pathway couples light to immediate-early gene expression in the suprachiasmatic nucleus. Eur J Neurosci, 17(8), 1617-1627.

Ebihara, S., & Tsuji, K. (1980). Entrainment of the circadian activity rhythm to the light cycle: effective light intensity for a Zeitgeber in the retinal degenerate C3H mouse and the normal C57BL mouse. Physiol Behav, 24(3), 523-527.

Emery, P., & Reppert, S. M. (2004). A rhythmic Ror. Neuron, 43(4), 443-446.

Ercolini, A. , & Berti, R. (1975). Light sensitivity experiments and morphology studies of the blind phreatic fish *Phreatichthys andruzzii* Vinciguerra from Somalia. Monit Zool Ital (N.S) Suppl, 6, 29-43.

Escobar, C., Cailotto, C., Angeles-Castellanos, M., Delgado, R. S., & Buijs, R. M. (2009). Peripheral oscillators: the driving force for food-anticipatory activity. Eur J Neurosci, 30(9), 1665-1675.

Etchegaray, J. P., Lee, C., Wade, P. A., & Reppert, S. M. (2003). Rhythmic histone acetylation underlies transcription in the mammalian circadian clock. Nature, 421(6919), 177-182.

Etchegaray, J. P., Yang, X., DeBruyne, J. P., Peters, A. H., Weaver, D. R., Jenuwein, T., & Reppert, S. M. (2006). The polycomb group protein EZH2 is required for mammalian circadian clock function. J Biol Chem, 281(30), 21209-21215.

Falcon, J. (1999). Cellular circadian clocks in the pineal. Prog Neurobiol, 58(2), 121-162.

Falcon, J., Besseau, L., Sauzet, S., & Boeuf, G. (2007). Melatonin effects on the hypothalamopituitary axis in fish. Trends Endocrinol Metab, 18(2), 81-88.

Feng, D., Liu, T., Sun, Z., Bugge, A., Mullican, S. E., Alenghat, T., Liu, X. S., & Lazar, M. A. (2011). A circadian rhythm orchestrated by histone deacetylase 3 controls hepatic lipid metabolism. Science, 331(6022), 1315-1319.

Flores, M. V., Hall, C., Jury, A., Crosier, K., & Crosier, P. (2007). The zebrafish retinoid-related orphan receptor (*ror*) gene family. Gene Expr Patterns, 7(5), 535-543.

Foster, R. G. (1998). Shedding light on the biological clock. Neuron, 20(5), 829-832.

Freedman, M. S., Lucas, R. J., Soni, B., von Schantz, M., Munoz, M., David-Gray, Z., & Foster, R. (1999). Regulation of mammalian circadian behavior by non-rod, non-cone, ocular photoreceptors. Science, 284(5413), 502-504.

Gachon, F. (2007). Physiological function of PARbZip circadian clock-controlled transcription factors. Ann Med, 39(8), 562-571.

Ganguly, S., Mummaneni, P., Steinbach, P. J., Klein, D. C., & Coon, S. L. (2001). Characterization of the *Saccharomyces cerevisiae* homolog of the melatonin rhythm enzyme arylalkylamine N-acetyltransferase (EC 2.3.1.87). J Biol Chem, 276(50), 4723947247.

Gavriouchkina, D., Fischer, S., Ivacevic, T., Stolte, J., Benes, V., & Dekens, M. P. (2010). Thyrotroph embryonic factor regulates light-induced transcription of repair genes in zebrafish embryonic cells. PLoS One, 5(9), e12542.

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.

Glaser, F. T., & Stanewsky, R. (2005). Temperature synchronization of the *Drosophila* circadian clock. Curr Biol, 15(15), 1352-1363.

Glaser, F. T., & Stanewsky, R. (2007). Synchronization of the *Drosophila* circadian clock by temperature cycles. Cold Spring Harb Symp Quant Biol, 72, 233-242.

Gnerre, S., Maccallum, I., Przybylski, D., Ribeiro, F. J., Burton, J. N., Walker, B. J., Sharpe, T., Hall, G., Shea, T. P., Sykes, S., et al. (2010). High-quality draft assemblies of mammalian genomes from massively parallel sequence data. Proc Natl Acad Sci USA, 108(4), 1513-1518.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnol, 29(7), 644–652. doi: 10.1038/nbt.1883.

Gross, J. B., Protas, M., Conrad, M., Scheid, P. E., Vidal, O., Jeffery, W. R., Borowsky, R., & Tabin, C. J. (2008). Synteny and candidate gene prediction using an anchored linkage map of *Astyanax mexicanus*. Proc Natl Acad Sci U S A, 105(51), 20106-20111.

Gross, J. B., Furterer, A., Carlson, B. M., Stahl, B. A. (2013) An Integrated Transcriptome-Wide Analysis of Cave and Surface Dwelling *Astyanax mexicanus*. PLoSONE 8(2): e55659. doi:10.1371/journal.pone.0055659

Guler, A. D., Ecker, J. L., Lall, G. S., Haq, S., Altimus, C. M., Liao, H. W., Barnard, A. R., Cahill, H., Badea, T. C., Zhao, H., Hankins, M. W., Berson, D. M., Lucas, R. J., Yau, K. W., & Hattar, S. (2008). Melanopsin cells are the principal conduits for rod-cone input to non-image-forming vision. Nature, 453(7191), 102-105.

Gunn, J. (2004). Encyclopedia of Caves and Karst Science. New York, Fitzroy Dearborn, 902 p.

Guo, A. Y., Zhu, Q. H., Chen, X., Luo, J. C. (2007). GSDS: a gene structure display server. Yi Chuan, 29(8), 1023-6.

Haffter, P., Granato, M., Brand, M., Mullins, M. C., Hammerschmidt, M., Kane, D. A., Odenthal, J., van Eeden, F. J., Jiang, Y. J., Heisenberg, C. P., Kelsh, R. N., Furutani-Seiki, M., Vogelsang, E., Beuchle, D., Schach, U., Fabian, C., & Nusslein-Volhard, C. (1996). The identification of genes with unique and essential functions in the development of the zebrafish, *Danio rerio*. Development, 123, 1-36.

Hannibal, J., Hindersson, P., Ostergaard, J., Georg, B., Heegaard, S., Larsen, P. J., & Fahrenkrug, J. (2004). Melanopsin is expressed in PACAP-containing retinal ganglion cells of the human retinohypothalamic tract. Invest Ophthalmol Vis Sci, 45(11), 42024209.

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

Hatori, M., Le, H., Vollmers, C., Keding, S. R., Tanaka, N., Buch, T., Waisman, A., Schmedt, C., Jegla, T., & Panda, S. (2008). Inducible ablation of melanopsin-expressing retinal ganglion cells reveals their central role in non-image forming visual responses. PLoS One, 3(6), e2451.

Hattar, S., Liao, H. W., Takao, M., Berson, D. M., & Yau, K. W. (2002). Melanopsin containing retinal ganglion cells: architecture, projections, and intrinsic photosensitivity. Science, 295(5557), 1065-1070.

Hattar, S., Lucas, R. J., Mrosovsky, N., Thompson, S., Douglas, R. H., Hankins, M. W., Lem, J., Biel, M., Hofmann, F., Foster, R. G., & Yau, K. W. (2003). Melanopsin and rod-cone photoreceptive systems account for all major accessory visual functions in mice. Nature, 424(6944), 76-81.

Hawes, R. S. (1945). On the eyes and reactions to light of *Proteus anguinus*. Q J Microsc Sci, 86, 1-53.

Hecht, M. K., Wallace, B., & Wilkens, H. (1988). Evolution and Genetics of Epigean and Cave *Astyanax fascistus* (Characidar, Pisces) Support for the Neutral Mutation Theory. Evolutionary Biology, 23, 271-367.

Hinaux, H., Poulain, J., Da Silva, C., Noirot, C., Jeffery, W. R., et al. (2013) *De Novo* Sequencing of *Astyanax mexicanus* Surface Fish and Pachòn Cavefish Transcriptomes Reveals Enrichment of Mutations in Cavefish Putative Eye Genes. PLoS ONE 8(1): e53553. doi:10.1371/journal.pone.0053553

Hirayama, J., Cho, S., & Sassone-Corsi, P. (2007). Circadian control by the reduction/oxidation pathway: catalase represses light-dependent clock gene expression in the zebrafish. Proc Natl Acad Sci U S A, 104(40), 15747-15752.

Hockberger, P. E., Skimina, T. A., Centonze, V. E., Lavin, C., Chu, S., Dadras, S., Reddy, J. K., & White, J. G. (1999). Activation of flavin-containing oxidases underlies light-induced production of H_2O_2 in mammalian cells. Proc Natl Acad Sci U S A, 96(11), 6255-6260.

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. Proc Natl Acad Sci U S A, 95(10), 5474-5479.

Howe, K., Clark, M. D., Torroja, C. F., Torrance, J., Berthelot, C., et al. (2013). The zebrafish reference genome sequence and its relationship to the human genome. Nature, 496, 498-503.

Hurd, M. W., Debruyne, J., Straume, M., & Cahill, G. M. (1998). Circadian rhythms of locomotor activity in zebrafish. Physiol Behav, 65(3), 465-472.

Ingham, P. W. (2009). The power of the zebrafish for disease analysis. Hum Mol Genet, 18(R1), R107-112.

Jeffery, W. R. (2001). Cavefish as a model system in evolutionary developmental biology. Dev Biol, 231(1), 1-12.

Jeffery, W. R. (2009). Regressive evolution in *Astyanax* cavefish. Annu Rev Genet, 43, 25-47.

Kakizawa, T., Nishio, S., Triqueneaux, G., Bertrand, S., Rambaud, J., & Laudet, V. (2007). Two differentially active alternative promoters control the expression of the zebrafish orphan nuclear receptor gene *Rev-erb alpha*. J Mol Endocrinol, 38(5), 555-568.

Kaneko, M., & Cahill, G. M. (2005). Light-dependent development of circadian gene expression in transgenic zebrafish. PLoS Biol, 3(2), e34.

Kaneko, M., Hernandez-Borsetti, N., & Cahill, G. M. (2006). Diversity of zebrafish peripheral oscillators revealed by luciferase reporting. Proc Natl Acad Sci U S A, 103(39), 1461414619.

Katada, S., & Sassone-Corsi, P. (2010). The histone methyltransferase MLL1 permits the oscillation of circadian gene expression. Nat Struct Mol Biol, 17(12), 1414-1421.

Kelley, J. L., Passow, C. N., Plath, M., Arias Rodriguez, L., Yee, M. C., Tobler, M. (2012). Genomic resources for a model in adaptation and speciation research: characterization of the *Poecilia mexicana* transcriptome. BMC Genomics, 13:652

King, D. P., Vitaterna, M. H., Chang, A. M., Dove, W. F., Pinto, L. H., Turek, F. W., &

Takahashi, J. S. (1997). The mouse *Clock* mutation behaves as an antimorph and maps within the W19H deletion, distal of Kit. Genetics, 146(3), 1049-1060.

Klein, D.M., & Reppert, RY. (1991). Suprachiasmatic nucleus -The Mind´s Clock. New York: Oxford University Press.

Kobayashi, Y., Ishikawa, T., Hirayama, J., Daiyasu, H., Kanai, S., Toh, H., Fukuda, I., Tsujimura, T., Terada, N., Kamei, Y., Yuba, S., Iwai, S., & Todo, T. (2000). Molecular analysis of zebrafish photolyase/cryptochrome family: two types of cryptochromes present in zebrafish. Genes Cells, 5(9), 725-738.

Koike, N., Yoo, S. H., Huang, H. C., Kumar, V., Lee, C., Kim, T. K., & Takahashi, J. S. (2012). Transcriptional architecture and chromatin landscape of the core circadian clock in mammals. Science, 338(6105), 349-354.

Konopka, R. J., & Benzer, S. (1971). Clock mutants of *Drosophila melanogaster*. Proc Natl Acad Sci U S A, 68(9), 2112-2116.

Korf, H. W., Schomerus, C., & Stehle, J. H. (1998). The pineal organ, its hormone melatonin, and the photoneuroendocrine system. Adv Anat Embryol Cell Biol, 146, 1-100.

Kruckenhauser, L., Haring, E., Seemann, R., & Sattmann, H. (2011). Genetic differentiation between cave and surface-dwelling populations of *Garra barreimiae* (Cyprinidae) in Oman. BMC Evol Biol, 11, 172.

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.

Kwok, C., Korn, R. M., Davis, M. E., Burt, D. W., Critcher, R., McCarthy, L., Paw, B. H., Zon, L. I., Goodfellow, P. N., & Schmitt, K. (1998). Characterization of whole genome radiation hybrid mapping resources for non-mammalian vertebrates. Nucleic Acids Res, 26(15), 3562-3566.

Lahiri, K., Vallone, D., Gondi, S. B., Santoriello, C., Dickmeis, T., & Foulkes, N. S. (2005). Temperature regulates transcription in the zebrafish circadian clock. PLoS Biol, 3(11), e351.

Lee, Y., & Montell, C. (2013). *Drosophila* TRPA1 Functions in Temperature Control of Circadian Rhythm in Pacemaker Neurons. J Neurosci, 33(16), 6716-6725.

Levine, J. D., Funes, P., Dowse, H. B., & Hall, J. C. (2002). Resetting the circadian clock by social experience in *Drosophila melanogaster*. Science, 298(5600), 2010-2012.

Li, P., Chaurasia, S. S., Gao, Y., Carr, A. L., Iuvone, P. M., & Li, L. (2008). CLOCK is required for maintaining the circadian rhythms of Opsin mRNA expression in photoreceptor cells. J Biol Chem, 283(46), 31673-31678.

Liu, C., & Reppert, S. M. (2000). GABA synchronizes clock cells within the suprachiasmatic circadian clock. Neuron, 25(1), 123-128.

Lopez-Olmeda, J. F., & Sanchez-Vazquez, F. J. (2009). Zebrafish temperature selection and synchronization of locomotor activity circadian rhythm to ahemeral cycles of light and temperature. Chronobiol Int, 26(2), 200-218.

Lopez-Olmeda, J. F., & Sanchez-Vazquez, F. J. (2010). Feeding rhythms in fish: from behavioral to molecular approach. In: Kulczykowska E, Popek W, Kapoor BG, eds. Biological clocks in fish. EnfieldNH: Science Publishers., pp 155-184.

Lopez-Olmeda, J. F., Tartaglione, E. V., de la Iglesia, H. O., & Sanchez-Vazquez, F. J. (2010). Feeding entrainment of food-anticipatory activity and *per1* expression in the brain and liver of zebrafish under different lighting and feeding conditions. Chronobiol Int, 27(7), 1380-1400.

Lucas, R. J., Freedman, M. S., Munoz, M., Garcia-Fernandez, J. M., & Foster, R. G. (1999). Regulation of the mammalian pineal by non-rod, non-cone, ocular photoreceptors. Science, 284(5413), 505-507.

Lucas, R. J., Hattar, S., Takao, M., Berson, D. M., Foster, R. G., & Yau, K. W. (2003).

Diminished pupillary light reflex at high irradiances in melanopsin-knockout mice. Science, 299(5604), 245-247.

Maier, B., Wendt, S., Vanselow, J. T., Wallach, T., Reischl, S., Oehmke, S., Schlosser, A., & Kramer, A. (2009). A large-scale functional RNAi screen reveals a role for CK2 in the mammalian circadian clock. Genes Dev, 23(6), 708-718.

Mano, H., Kojima, D., & Fukada, Y. (1999). Exo-rhodopsin: a novel rhodopsin expressed in the zebrafish pineal gland. Brain Res Mol Brain Res, 73(1-2), 110-118.

Masri, S., & Sassone-Corsi, P. (2010). Plasticity and specificity of the circadian epigenome. Nat Neurosci, 13(11), 1324-1329.

McGaugh, S. E., Gross, J. B., Aken, B., Blin, M., Borowsky, R., Chalopin, D., Hinaux, H., Jeffery, W. R., Keene, A., Ma, L., Minx, P., Murphy, D., O'Quin, K. E., Rétaux, S., Rohner, N., Searle, S. M. J., Stahl, B. A., Tabin, C., Volff, J., Yoshizawa, M., Warren, W. C. (2014). The cavefish genome reveals candidate genes for eye loss. Nature Communications. 5(5307), DOI: 10.1038/ncomms6307.

Mehra, A., Baker, C. L., Loros, J. J., & Dunlap, J. C. (2009). Post-translational modifications in circadian rhythms. Trends Biochem Sci, 34(10), 483-490.

Menaker, M., Moreira, L. F., & Tosini, G. (1997). Evolution of circadian organization in vertebrates. Braz J Med Biol Res, 30(3), 305-313.

Menaker, M., Takahashi, J. S., & Eskin, A. (1978). The physiology of circadian pacemakers. Annu Rev Physiol, 40, 501-526.

Meyer, A., & Schartl, M. (1999). Gene and genome duplications in vertebrates: the one-to-four (-to-eight in fish) rule and the evolution of novel gene functions. Curr Opin Cell Biol, 11(6), 699-704.

Moore, R.Y. , & Eichler, V.B. (1972). Loss of a circadian adrenal corticosterone rhythm following suprachiasmatic lesions in the rat. Brain Res., 42, 201-206.

Mori, T., Binder, B., & Johnson, C. H. (1996). Circadian gating of cell division in cyanobacteria growing with average doubling times of less than 24 hours. Proc Natl Acad Sci U S A, 93(19), 10183-10188.

Moutsaki, P., Whitmore, D., Bellingham, J., Sakamoto, K., David-Gray, Z. K., & Foster, R. G. (2003). Teleost multiple tissue (*tmt*) opsin: a candidate photopigment regulating the peripheral clocks of zebrafish? Brain Res Mol Brain Res, 112(1-2), 135-145.

Mracek, P., Santoriello, C., Idda, M. L., Pagano, C., Ben-Moshe, Z., Gothilf, Y., Vallone, D., & Foulkes, N. S. (2012). Regulation of *per* and *cry* genes reveals a central role for the D-box enhancer in light-dependent gene expression. PLoS One, 7(12), e51278.

Mullins, M. C., & Nusslein-Volhard, C. (1993). Mutational approaches to studying embryonic pattern formation in the zebrafish. Curr Opin Genet Dev, 3(4), 648-654.

Nagoshi, E., Saini, C., Bauer, C., Laroche, T., Naef, F., & Schibler, U. (2004). Circadian gene expression in individual fibroblasts: cell-autonomous and self-sustained oscillators pass time to daughter cells. Cell, 119(5), 693-705.

Nakahata, Y., Grimaldi, B., Sahar, S., Hirayama, J., & Sassone-Corsi, P. (2007). Signaling to the circadian clock: plasticity by chromatin remodeling. Curr Opin Cell Biol, 19(2), 230-237.

Nawathean, P., & Rosbash, M. (2004). The doubletime and CKII kinases collaborate to potentiate *Drosophila* PER transcriptional repressor activity. Mol Cell, 13(2), 213-223.

Noche, Lu, Goldstein-Kral, Glasgow, Liang. (2011). Circadian rhythms in the pineal organ persist in zebrafish larvae that lack ventral brain. BMC Neurosci., 13(12:7).

Ozturk, N., Selby, C. P., Song, S. H., Ye, R., Tan, C., Kao, Y. T., Zhong, D., & Sancar, A. (2009). Comparative photochemistry of animal type 1 and type 4 cryptochromes. Biochemistry, 48(36), 8585-8593.

Panda, S., Nayak, S. K., Campo, B., Walker, J. R., Hogenesch, J. B., & Jegla, T. (2005). Illumination of the melanopsin signaling pathway. Science, 307(5709), 600-604.

Panda, S., Provencio, I., Tu, D. C., Pires, S. S., Rollag, M. D., Castrucci, A. M., Pletcher, M. T., Sato, T. K., Wiltshire, T., Andahazy, M., Kay, S. A., Van Gelder, R. N., & Hogenesch, J. B. (2003). Melanopsin is required for non-image-forming photic responses in blind mice. Science, 301(5632), 525-527.

Pando, M. P., Morse, D., Cermakian, N., & Sassone-Corsi, P. (2002). Phenotypic rescue of a peripheral clock genetic defect via SCN hierarchical dominance. Cell, 110(1), 107-117.

Pando, M. P., Pinchak, A. B., Cermakian, N., & Sassone-Corsi, P. (2001). A cell-based system that recapitulates the dynamic light-dependent regulation of the vertebrate clock. Proc Natl Acad Sci U S A, 98(18), 10178-10183.

Pando, M. P., & Sassone-Corsi, P. (2002). Unraveling the mechanisms of the vertebrate circadian clock: zebrafish may light the way. Bioessays, 24(5), 419-426.

Pando, Pinchak, Cermakian, Sassone-Corsi. (2001). A cell-based system that recapitulates the dynamic light-dependent regulation of the vertebrate clock. Proc Natl Acad Sci U S A, 28(98(18)), 10178-10183.

Peirson, S. N., Halford, S., & Foster, R. G. (2009). The evolution of irradiance detection: melanopsin and the non-visual opsins. Philos Trans R Soc Lond B Biol Sci, 364(1531), 2849-2865.

Pevet, P., & Challet, E. (2011). Melatonin: both master clock output and internal time-giver in the circadian clocks network. J Physiol Paris, 105(4-6), 170-182.

Pierce, L. X., Noche, R. R., Ponomareva, O., Chang, C., & Liang, J. O. (2008). Novel functions for *Period 3* and *Exo-rhodopsin* in rhythmic transcription and melatonin biosynthesis within the zebrafish pineal organ. Brain Res, 1223, 11-24.

Pipan, T , & Culver, DC. (2012). Convergence and divergence in the subterranean realm: a reassessment. Biological Journal of the Linnean Society, 107(1-14).

Pittendrigh, C. S. (1960). Circadian rhythms and the circadian organization of living systems. Cold Spring Harb Symp Quant Biol, 25, 159-184.

Pittendrigh, C. S. (1967). Circadian systems. I. The driving oscillation and its assay in Drosophila pseudoobscura. Proc Natl Acad Sci U S A, 58(4), 1762-1767.

Pittendrigh, C. S. (1993). Temporal organization: reflections of a Darwinian clock-watcher. Annu Rev Physiol, 55, 16-54.

Plautz, J. D., Kaneko, M., Hall, J. C., & Kay, S. A. (1997). Independent photoreceptive circadian clocks throughout *Drosophila*. Science, 278(5343), 1632-1635.

Ponting, C. P., & Aravind, L. (1997). PAS: a multifunctional domain family comes to light. Curr Biol, 7(11), R674-677.

Postlethwait, J. H., Yan, Y. L., Gates, M. A., Horne, S., Amores, A., Brownlie, A., Donovan, A., Egan, E. S., Force, A., Gong, Z., Goutel, C., Fritz, A., Kelsh, R., Knapik, E., Liao, E., Paw, B., Ransom, D., Singer, A., Thomson, M., Abduljabbar, T. S., Yelick, P., Beier, D., Joly, J. S., Larhammar, D., Rosa, F., Westerfield, M., Zon, L. I., Johnson, S. L., & Talbot, W. S. (1998). Vertebrate genome evolution and the zebrafish gene map. Nat Genet, 18(4), 345-349.

Powell, S., Szklarczyk, D., Trachana, K., Roth, A., Kuhn, M., Muller, J., Arnold, R., Rattei, T., Letunic, I., Doerks, T., et al. (2011). eggNOG v3.0: orthologous groups covering 1133 organisms at 41 different taxonomic ranges. Nucleic Acids Res, 40, D284-D289.

Preitner, N., Damiola, F., Lopez-Molina, L., Zakany, J., Duboule, D., Albrecht, U., & Schibler, U. (2002). The orphan nuclear receptor REV-ERBalpha controls circadian transcription within the positive limb of the mammalian circadian oscillator. Cell, 110(2), 251-260.

Protas, M. E., Hersey, C., Kochanek, D., Zhou, Y., Wilkens, H., Jeffery, W. R., Zon, L. I., Borowsky, R., & Tabin, C. J. (2006). Genetic analysis of cavefish reveals molecular convergence in the evolution of albinism. Nat Genet, 38(1), 107-111.

Protas, M., & Jeffery, W. R. (2012). Evolution and development in cave animals: from fish to crustaceans. Wiley Interdiscip Rev Dev Biol, 1(6), 823-845.

Proudlove, G. . (2006). Subterranean Fishes of the World. International Society for Subterranean Biology.

Provencio, I., Jiang, G., De Grip, W. J., Hayes, W. P., & Rollag, M. D. (1998). Melanopsin: An opsin in melanophores, brain, and eye. Proc Natl Acad Sci U S A, 95(1), 340-345.

Provencio, I., Rodriguez, I. R., Jiang, G., Hayes, W. P., Moreira, E. F., & Rollag, M. D. (2000). A novel human opsin in the inner retina. J Neurosci, 20(2), 600-605.

Raghuram, S., Stayrook, K. R., Huang, P., Rogers, P. M., Nosie, A. K., McClure, D. B., Burris, L. L., Khorasanizadeh, S., Burris, T. P., & Rastinejad, F. (2007). Identification of heme as the ligand for the orphan nuclear receptors REV-ERBalpha and REV-ERBbeta. Nat Struct Mol Biol, 14(12), 1207-1213.

Ralph, M. R., Foster, R. G., Davis, F. C., & Menaker, M. (1990). Transplanted suprachiasmatic nucleus determines circadian period. Science, 247(4945), 975-978.

Ramsey, I. S., Delling, M., & Clapham, D. E. (2006). An introduction to TRP channels. Annu Rev Physiol, 68, 619-647.

Reddy, P., Zehring, W. A., Wheeler, D. A., Pirrotta, V., Hadfield, C., Hall, J. C., & Rosbash, M. (1984). Molecular analysis of the *period* locus in *Drosophila melanogaster* and identification of a transcript involved in biological rhythms. Cell, 38(3), 701-710.

Refinetti, R. (2010). Entrainment of circadian rhythm by ambient temperature cycles in mice. J Biol Rhythms, 25(4), 247-256.

Reppert, S. M., & Weaver, D. R. (2001). Molecular analysis of mammalian circadian rhythms. Annu Rev Physiol, 63, 647-676.

Roenneberg, T., Daan, S., & Merrow, M. (2003). The art of entrainment. J Biol Rhythms, 18(3), 183-194.

Roenneberg, T., & Foster, R. G. (1997). Twilight times: light and the circadian system. Photochem Photobiol, 66(5), 549-561.

Romero, A., Green, S. M, Romero, A. , Lelonek, M. M., & Stropnicky, K. C. (2003). One eye but no vision: Cavefish with induced eyes do not respond to light. Journal of Experimental Zoology Part B: Molecular and Developmental Evolution, 300B(1), 72-79.

Ruby, N. F., Brennan, T. J., Xie, X., Cao, V., Franken, P., Heller, H. C., & O'Hara, B. F. (2002). Role of melanopsin in circadian responses to light. Science, 298(5601), 22112213.

Sadoglu, P. (1957). Mendelian inheritance in hybrids between the Mexican blind fish and their overground ancestors. Verh Dtsch Zool Ges., 432-439.

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), 527-537.

Schibler, U. (2007). The daily timing of gene expression and physiology in mammals. Dialogues Clin Neurosci, 9(3), 257-272.

Schibler, U., & Sassone-Corsi, P. (2002). A web of circadian pacemakers. Cell, 111(7), 919-922.

Sehadova, H., Glaser, F. T., Gentile, C., Simoni, A., Giesecke, A., Albert, J. T., & Stanewsky, R. (2009). Temperature entrainment of *Drosophila*'s circadian clock involves the gene *nocte* and signaling from peripheral sensory tissues to the brain. Neuron, 64(2), 251-266.

Sexton, T., Buhr, E., & Van Gelder, R. N. (2012). Melanopsin and mechanisms of non-visual ocular photoreception. J Biol Chem, 287(3), 1649-1656.

Shearman, L. P., Sriram, S., Weaver, D. R., Maywood, E. S., Chaves, I., Zheng, B., Kume, K., Lee, C. C., van der Horst, G. T., Hastings, M. H., & Reppert, S. M. (2000). Interacting molecular loops in the mammalian circadian clock. Science, 288(5468), 1013-1019.

Shearman, L. P., Zylka, M. J., Reppert, S. M., & Weaver, D. R. (1999). Expression of basic helix-loop-helix/PAS genes in the mouse suprachiasmatic nucleus. Neuroscience, 89(2), 387-397.

Shearman, L. P., Zylka, M. J., Weaver, D. R., Kolakowski, L. F., Jr., & Reppert, S. M. (1997). Two period homologs: circadian expression and photic regulation in the suprachiasmatic nuclei. Neuron, 19(6), 1261-1269.

Shigeyoshi, Y., Taguchi, K., Yamamoto, S., Takekida, S., Yan, L., Tei, H., Moriya, T., Shibata, S., Loros, J. J., Dunlap, J. C., & Okamura, H. (1997). Light-induced resetting of a mammalian circadian clock is associated with rapid induction of the m*Per1* transcript. Cell, 91(7), 1043-1053.

Solt, L. A., & Burris, T. P. (2012). Action of RORs and their ligands in (patho)physiology. Trends Endocrinol Metab, 23(12), 619-627.

Stephan, F. K. (2002). The "other" circadian system: food as a Zeitgeber. J Biol Rhythms, 17(4), 284-292.

Storch, K. F., & Weitz, C. J. (2009). Daily rhythms of food-anticipatory behavioral activity do not require the known circadian clock. Proc Natl Acad Sci U S A, 106(16), 6808-6813.

Strecker, U., Hausdorf, B., & Wilkens, H. (2012). Parallel speciation in *Astyanax* cave fish (Teleostei) in Northern Mexico. Mol Phylogenet Evol, 62(1), 62-70.

Takahashi, J. S., Hamm, H., & Menaker, M. (1980). Circadian rhythms of melatonin release from individual superfused chicken pineal glands in vitro. Proc Natl Acad Sci U S A, 77(4), 2319-2322.

Tamai, T. K., Vardhanabhuti, V., Arthur, S., Foulkes, N. S., & Whitmore, D. (2003). Flies and fish: birds of a feather. J Neuroendocrinol, 15(4), 344-349.

Tarttelin, E. E., Frigato, E., Bellingham, J., Di Rosa, V., Berti, R., Foulkes, N. S., Lucas, R. J., & Bertolucci, C. (2012). Encephalic photoreception and phototactic response in the troglobiont Somalian blind cavefish *Phreatichthys andruzzii*. J Exp Biol, 215(Pt 16), 2898-2903.

Tatusov, R. L., Fedorova, N. D., Jackson, J. D., Jacobs, A. R., Kiryutin, B., Koonin, E. V., et al. (2003). The COG database: an updated version includes eukaryotes. BMC Bioinformatics, 4(41), doi: 10.1186/1471-2105-4-41.

Tei, H., Okamura, H., Shigeyoshi, Y., Fukuhara, C., Ozawa, R., Hirose, M., & Sakaki, Y. (1997). Circadian oscillation of a mammalian homologue of the *Drosophila period* gene. Nature, 389(6650), 512-516.

Timmermann, M., & Plath, M. (2009). Phototactic response and light sensitivity in an epigean and a hypogean population of a barb (*Garra barreimiae*, Cyprinidae). Aquatic Ecology, 43, 539-547.

Toh, K. L., Jones, C. R., He, Y., Eide, E. J., Hinz, W. A., Virshup, D. M., Ptacek, L. J., & Fu, Y. H. (2001). An h*Per2* phosphorylation site mutation in familial advanced sleep phase syndrome. Science, 291(5506), 1040-1043.

Tomioka, K., & Matsumoto, A. (2010). A comparative view of insect circadian clock systems. Cell Mol Life Sci, 67(9), 1397-1406.

Tomita, T., Miyazaki, K., Onishi, Y., Honda, S., Ishida, N., & Oishi, K. (2010). Conserved amino acid residues in C-terminus of PERIOD 2 are involved in interaction with CRYPTOCHROME 1. Biochim Biophys Acta, 1803(4), 492-498.

Tosini, G., & Menaker, M. (1996). Circadian rhythms in cultured mammalian retina. Science, 272(5260), 419-421.

Trontelj, P., Blejec, A., & Fiser, C. (2012). Ecomorphological convergence of cave communities. Evolution, 66(12), 3852-3865.

Tsuchiya, Y., Akashi, M., & Nishida, E. (2003). Temperature compensation and temperature resetting of circadian rhythms in mammalian cultured fibroblasts. Genes Cells, 8(8), 713-720.

Uchida, Y., Hirayama, J., & Nishina, H. (2010). A common origin: signaling similarities in the regulation of the circadian clock and DNA damage responses. Biol Pharm Bull, 33(4), 535-544.

Underwood, H., & Groos, G. (1982). Vertebrate circadian rhythms: retinal and extraretinal photoreception. Experientia, 38(9), 1013-1021.

Vallone, D., Gondi, S. B., Whitmore, D., & Foulkes, N. S. (2004). E-box function in a *period* gene repressed by light. Proc Natl Acad Sci U S A, 101(12), 4106-4111.

Vallone, D., Lahiri, K., Dickmeis, T., & Foulkes, N. S. (2007). Start the clock! Circadian rhythms and development. Dev Dyn, 236(1), 142-155.

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), 627-630.

Van Ooijen, G., & Millar, A. J. (2012). Non-transcriptional oscillators in circadian timekeeping. Trends Biochem Sci, 37(11), 484-492.

Varatharasan, N., Croll, R. P., & Franz-Odendaal, T. (2009). Taste bud development and patterning in sighted and blind morphs of *Astyanax mexicanus*. Dev Dyn, 238(12), 30563064.

Vatine, G., Vallone, D., Appelbaum, L., Mracek, P., Ben-Moshe, Z., Lahiri, K., Gothilf, Y., & Foulkes, N. S. (2009). Light directs zebrafish *Period2* expression via conserved D and E boxes. PLoS Biol, 7(10), e1000223.

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), 719-725.

Vitaterna, M. H., Selby, C. P., Todo, T., Niwa, H., Thompson, C., Fruechte, E. M., Hitomi, K., Thresher, R. J., Ishikawa, T., Miyazaki, J., Takahashi, J. S., & Sancar, A. (1999). Differential regulation of mammalian period genes and circadian rhythmicity by cryptochromes 1 and 2. Proc Natl Acad Sci U S A, 96(21), 12114-12119.

Wager-Smith, K., & Kay, S. A. (2000). Circadian rhythm genetics: from flies to mice to humans. Nat Genet, 26(1), 23-27.

Wang, H. (2008). Comparative analysis of teleost fish genomes reveals preservation of different ancient clock duplicates in different fishes. Mar Genomics, 1(2), 69-78.

Weger, B. D., Sahinbas, M., Otto, G. W., Mracek, P., Armant, O., Dolle, D., Lahiri, K., Vallone, D., Ettwiller, L., Geisler, R., Foulkes, N. S., & Dickmeis, T. (2011). The light responsive transcriptome of the zebrafish: function and regulation. PLoS One, 6(2), e17080.

Wessel, A., Hoch, H., Asche, M., von Rintelen, T., Stelbrink, B., Heck, V., Stone, F. D., & Howarth, F. G. (2013). Founder effects initiated rapid species radiation in Hawaiian cave planthoppers. Proc Natl Acad Sci U S A, 110(23), 9391-9396.

Whitmore, D., Foulkes, N. S., & Sassone-Corsi, P. (2000). Light acts directly on organs and cells in culture to set the vertebrate circadian clock. Nature, 404(6773), 87-91.

Whitmore, D., Foulkes, N. S., Strahle, U., & Sassone-Corsi, P. (1998). Zebrafish Clock rhythmic expression reveals independent peripheral circadian oscillators. Nat Neurosci, 1(8), 701-707.

Wilhelm, F. M., Taylor, S. J. , & Adams, G. L. (2006). Comparison of routine metabolic rates of the stygobite, *Gammarus acherondytes* (Amphipoda: Gammaridae) and the stygophile, *Gammarus troglophilus*. Freshwater Biology, 51(6), 1162-1174.

Wilkens, H. (1971). Genetic interpretation of regressive evolutionary processes: studies of hybrid eyes of two *Astyanax* cave populations (Characidae, Pisces). Evolution, 25, 530–544.

Wilkens, H. (2001). Convergent Adaptations to Cave Life in the *Rhamdia Laticauda* Catfish Group (Pimelodidae, Teleostei). Environmental Biology of Fishes, 62(1-3), pp 251-261.

Wilkens, H. (2010). Genes, modules and the evolution of cave fish. Heredity (Edinb), 105(5), 413-422.

Wilkens, H. (2011). Variability and loss of functionless traits in cave animals. Reply to Jeffery (2010). Heredity (Edinb), 106(4), 707-708.

Wilkens, H., & Strecker, U. (2003). Convergent evolution of the cavefish *Astyanax* (Characidae, Teleostei): genetic evidence from reduced eye-size and pigmentation. Biological Journal of the Linnean Society, 80(4), 545-554.

Yamaguchi, S., Isejima, H., Matsuo, T., Okura, R., Yagita, K., Kobayashi, M., & Okamura, H. (2003). Synchronization of cellular clocks in the suprachiasmatic nucleus. Science, 302(5649), 1408-1412.

Yamamoto, Y. (2004). Cavefish. Curr Biol, 14(22), R943.

Yamamoto, Y., Byerly, M. S., Jackman, W. R., & Jeffery, W. R. (2009). Pleiotropic functions of embryonic sonic hedgehog expression link jaw and taste bud amplification with eye loss during cavefish evolution. Dev Biol, 330(1), 200-211.

Yamamoto, Y., & Jeffery, W. R. (2000). Central role for the lens in cave fish eye degeneration. Science, 289(5479), 631-633.

Yamazaki, S., Numano, R., Abe, M., Hida, A., Takahashi, R., Ueda, M., Block, G. D., Sakaki, Y., Menaker, M., & Tei, H. (2000). Resetting central and peripheral circadian oscillators in transgenic rats. Science, 288(5466), 682-685.

Yin, L., Wu, N., Curtin, J. C., Qatanani, M., Szwergold, N. R., Reid, R. A., Waitt, G. M., Parks, D. J., Pearce, K. H., Wisely, G. B., & Lazar, M. A. (2007). REV-ERBALPHA, a heme sensor that coordinates metabolic and circadian pathways. Science, 318(5857), 17861789.

Yokoyama, S. (2000). Molecular evolution of vertebrate visual pigments. Prog Retin Eye Res, 19(4), 385-419.

Yoo, S. H., Yamazaki, S., Lowrey, P. L., Shimomura, K., Ko, C. H., Buhr, E. D., Siepka, S. M., Hong, H. K., Oh, W. J., Yoo, O. J., Menaker, M., & Takahashi, J. S. (2004). PERIOD2::LUCIFERASE real-time reporting of circadian dynamics reveals persistent circadian oscillations in mouse peripheral tissues. Proc Natl Acad Sci U S A, 101(15), 5339-5346.

Yoshizawa, M., & Jeffery, W. R. (2011). Evolutionary tuning of an adaptive behavior requires enhancement of the neuromast sensory system. Commun Integr Biol, 4(1), 89-91.

Zehring, W. A., Wheeler, D. A., Reddy, P., Konopka, R. J., Kyriacou, C. P., Rosbash, M., & Hall, J. C. (1984). P-element transformation with *period* locus DNA restores rhythmicity to mutant, arrhythmic *Drosophila melanogaster*. Cell, 39(2 Pt 1), 369-376.

Zheng, B., Albrecht, U., Kaasik, K., Sage, M., Lu, W., Vaishnav, S., Li, Q., Sun, Z. S., Eichele, G., Bradley, A., & Lee, C. C. (2001). Non-redundant roles of the m*Per1* and m*Per2* genes in the mammalian circadian clock. Cell, 105(5), 683-694.

Ziv, L., Tovin, A., Strasser, D., & Gothilf, Y. (2007). Spectral sensitivity of melatonin suppression in the zebrafish pineal gland. Exp Eye Res, 84(1), 92-99.

Zylka, M. J., Shearman, L. P., Weaver, D. R., & Reppert, S. M. (1998). Three *period* homologs in mammals: differential light responses in the suprachiasmatic circadian clock and oscillating transcripts outside of brain. Neuron, 20(6), 1103-1110.

Acknowledgements

I would like first of all to thank my supervisor Prof. Cristiano Bertolucci who gave me the great opportunity to work in his laboratory. I want to thank him for his support and the patient guidance he has been giving through my entire PhD, starting from helpful discussions, seminar preparations and finally giving comments and correcting this thesis. I want to thank also Dr. Elena Frigato for her advices and for helping me to troubleshoot problems and teaching me many techniques at the beginning of my studies.

I would like to thank as well my coordinator Prof. Guido Barbujani for his help and friendly collaboration during my entire PhD at the University of Ferrara.

I am also very grateful to Prof. Dr. Nicholas S. Foulkes and Dr. Daniela Vallone who gave me the chance to work in their lab for most of my PhD, at the Karlsruhe Institute of Technology (KIT), Institute for Toxicology and Genetics (ITG). I want to thank them for their strong support, helpful supervision and the huge scientific knowledge I acquired from them during more than two years of collaboration. Big thanks also to our collaborator from the NGS Core Facility (KIT) Dr. Olivier Armant for his contribute to the Deep Sequencing analysis and for many helpful discussions and suggestions.

I also would like to thank Prof. Dr. Francisco Javier Sánchez-Vázquez from the University of Murcia (Spain) for the close collaboration and the useful periods of experiments performed under his guidance in Spain.

Last but not least I want to thank my wife and my family for all their strong support they have given me through this whole time, as well as the many friends from Italy, Germany and Spain who also supported and helped me in many different ways.