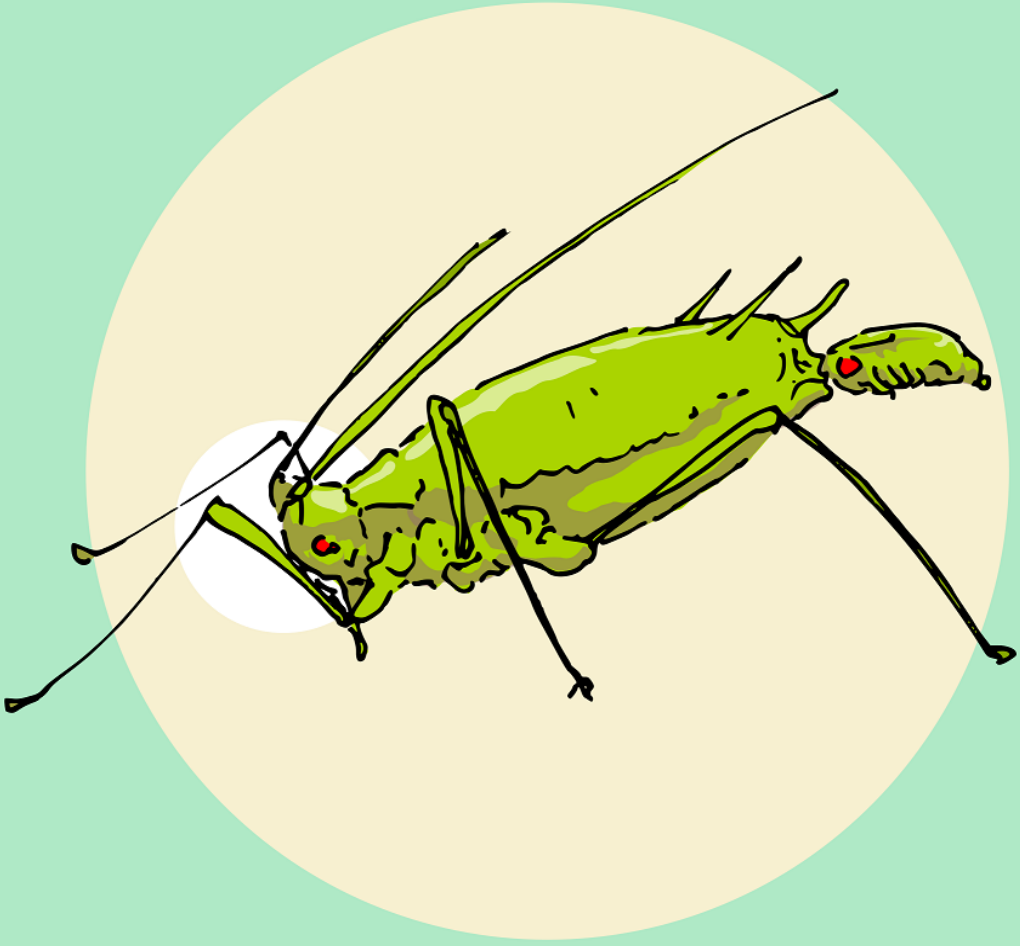
Transcriptional study of photoperiodic response in aphids and characterization of candidate photoperiodic photoreceptors

J. Mariano Collantes Alegre



VNIVERSITAT DE VALÈNCIA

PhD Thesis - Doctoral Programme 3101
Biodiversity and Evolutionary Biology (RD 99/2011)
Facultad de Ciencias Biológicas
Directed by Dr. David Martínez Torres. October 2020

Transcriptional study of photoperiodic response in aphids and characterization of candidate photoperiodic photoreceptors

J. Mariano Collantes Alegre

PhD Thesis

Doctoral Programme 3101 Biodiversity and Evolutionary Biology

(RD 99/2011)

Facultad de Ciencias Biológicas

Directed by Dr. David Martínez Torres

Valencia, Spain. October 2020

D. David Martínez Torres, Profesor Titular del Departamento de Genética y miembro del Instituto de Biología Integrativa de Sistemas de la Universitat de València

CERTIFICO: que Jorge Mariano Collantes Alegre, Licenciado en Ciencias Biológicas de la Universidad de Valencia, ha realizado el trabajo titulado:"Transcriptional study of photoperiodic response in aphids and characterization of candidate photoperiodic photoreceptor" bajo mi supervisión para optar al Grado de Doctor por la Universidad de Valencia.

Para que conste, en cumplimiento de la legislación vigente, firmo el presente informe en Valencia, a 15 de septiempre de 2020.

Firmado: Dr. David Martínez Torres

Agradecimientos

En estos años de tesis he recibido apoyo, ayuda y enseñanzas de muchas personas.

En primer lugar, me gustaría agradecer a mi director, David, por todo lo que he aprendido a su lado y por haber confiado en mi para llevar a cabo este proyecto.

A los investigadores y profesores del Instituto Cavanilles y del I2SysBio por su guía y consejos. También al personal del SCSIE de la UV (vivero, secuenciación y microscopía) que me ayudaron mucho más de lo que les correspondía y a Jorge por su paciencia a la hora de hacer cortes en el criostato.

Con especial cariño quiero agradecer a Quelo y Adrián, con quienes he compartido este largo recorrido en en las bancadas de laboratorio y fuera de ellas.

Quiero agradecer también a mis amigos Dani, Canxu y Jose; a los incansables Fernando, Manu y Rubén; a los compañeros y maestros desde la época en el Instituto Cavanilles: Sergio, Manzano, Florian,

Vanesa, Mariana, Carlos y a María. También a los grandes Paolo, Jesús y Emilio; a Nico, Alice y Víctor por ser fuentes inagotables de conocimiento sobre insectos.

A Ana González y a Rosa Martínez, cuyos pequeños consejos me resultaron de mucha ayuda. A todo MusSol Teatre por ser un cable a tierra. A la familia Monzó Beltrán por aguantarme y cuidarme.

Muy especialmente me gustaría agradecer a Lily, por su apoyo, cariño y compañía.

Por último, no tengo palabras suficientes para agradecer a mis padres, Jorge y Cristina y a mis hermanos Mati, Leo y Brenda. Gracias por su apoyo, confianza, consejo, cariño y por rodearme de sobrinos maravillosos.

Y ya que a juerza de golpes
La suerte nos dejó a flus
puede que allá veamos luz
y se acaben nuestras penas:
todas las tierras son güenas;
vamosnós, amigo Cruz.

José Hernández - "El gaucho Martín Fierro"

Contents

Chapter 1

General introduction

1.1 Aphid biology

Aphids (Hemiptera: Aphididae) are small pear-shaped insects present worldwide, but they are more frequently found and diverse in temperate zones of the northern hemisphere (Blackman & Eastop, 2000). There are around 5000 aphid species (Żyła et al., 2017), out of which 250 are considered agricultural pests (Oerke & Dehne, 1997), because they damage crops in many direct and indirect ways. Their mouthparts form a stylet used to pierce through the plant and feed on phloem sap. They cause direct damage to the plant by sucking its nutrients, but also through the toxicity of their saliva. Also, aphids serve as vectors for numerous plant viruses carried in their salivary glands. Moreover, aphids excrete honeydew, a sugar-rich

deposition where fungi develop and negatively affect the plant.

Aphids reproduce by cyclical parthenogenesis, i.e., they alternate one generation of individuals reproducing sexually with several generations of individuals reproducing parthenogenetically along the year (Figure 1.1). This alternation can lead to intricate life cycles. Both reproductive modes can differ in appearence, so there can exist different "morphs" of the same species. These different morphs can also live and feed on different host plant species. Heteroecious species spend both phases of their life cycle on different host plant species, while monoecious species can complete their whole life cycle on the same host. *Acyrthosiphon pisum* is a widely studied monoecious model species which can feed on different species of legumes (Fabaceae). During the favourable warm seasons, around 10-30 generations (Ogawa & Miura, 2014) of only parthenogenetic females are produced that give birth to their progeny through viviparity (see Figure 1.1). This all female progeny is produced from unfertilized eggs, thus being a case of "thelytoky". This clonal offspring are genetically identical to their mothers. Parthenogenetic morphs reproduce at high rates, thus generating dense populations and making aphids an even more damaging pest. Parthenogenetic females have a short generation time of about 10 days and exhibit telescopic generations, meaning that embryos inside parthenogenetic mothers start developing their own progeny even before being born.

As most aphid species, *A. pisum* goes through four nymphal or larval stages (L1 to L4) before reaching adulthood after the final molt (Figure 1.2).

In autumn, when days shorten and temperature decreases, *A. pisum* senses the difference in night length and the parthenogenetic females start producing sexual males (see Figure 1.3). These males are generated after a special meiosis which results in eggs with only one X chromosome. In this sex-determination system females have a chromosome set with two sexual chromosomes (XX), while males have only one (X0). Males are phenotypically different from females, with much smaller bodies and are often winged. The females that give birth to sexual individuals are called "sexuparae".

As autumn advances, the photoperiod keeps shortening and parthenogenetic females produce sexual females also known as oviparae, since, differently from the parthenogenetic ones, they lay eggs after mating with males. The eggs are in a state of dormancy, and hatch after approximately three months. These eggs are very resistant and can survive the harsh winter conditions. Hatching occurs in spring to start a new cycle by giving birth to the first generation of parthenogenetic females.

Since induction of sexual reproduction in aphids produces a dormant and resistant phase, it is usually considered as a diapause. Aphids, therefore, have a complex and interesting life cycle, during

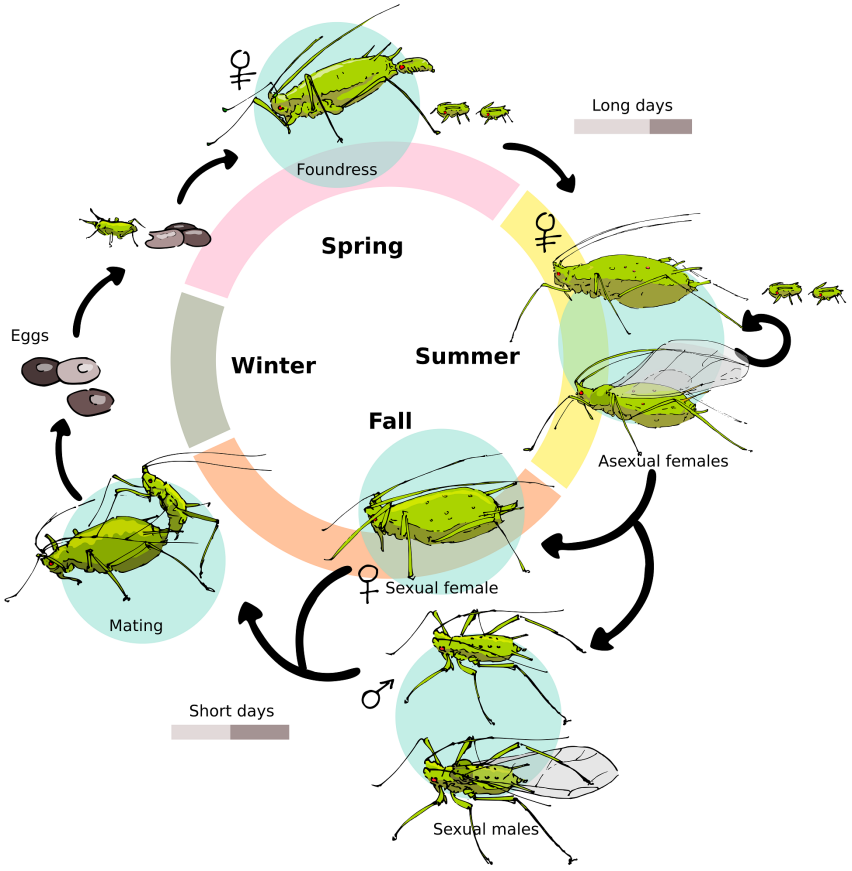


Figure 1.1: Life cycle of the pea aphid *Acyrthosiphon pisum*. During the favourable seasons (spring and summer) parthenogenetic viviparous females (i.e. virginoparae) give birth to several generations of clones. When days shorten (and nights are longer) with the arrival of autumn, parthenogenetic females alter the development of their embryos to start giving birth to males and, when winter is closer, to sexual oviparous females. Males and sexual females mate, and the latter lay overwintering eggs. These diapausing eggs survive the harsh winter conditions and hatch in spring, starting the cycle again. The pea aphid is a monoecious species, living and feeding on Fabaceae (laburnum, broom, peas, beans, etc.) all year long. Female sign with double stroke represents parthenogentic females. Colored bars represent long or short days, the clear portion of the bar indicates day length.

Figure 1.2: Adult parthenogenetic female of the pea aphid (*Acyrthosiphon pisum*) giving birth to a clonal daughter. The eyes of sister embryos are visible as dark red spots through the cuticle of the adult abdomen. Those embryos are already developing their own progeny. Taken from Barberà, 2017.

which genetically identical individuals display strikingly different phenotypes controlled by shortening of the photoperiod. The molecular pathways involved in the switch to sexual reproduction (i.e. in male or oviparae production) are poorly understood. While there might be two independent mechanisms for male and for sexual female production, it is also possible that one molecular cascade produces two different outcomes with slight variations.

Thus, there is a birth sequence coordinated with and induced by the sequential shortening of the photoperiod (MacKay, 1987; 1989). Although there is variability among strains (Mackay, 1989),

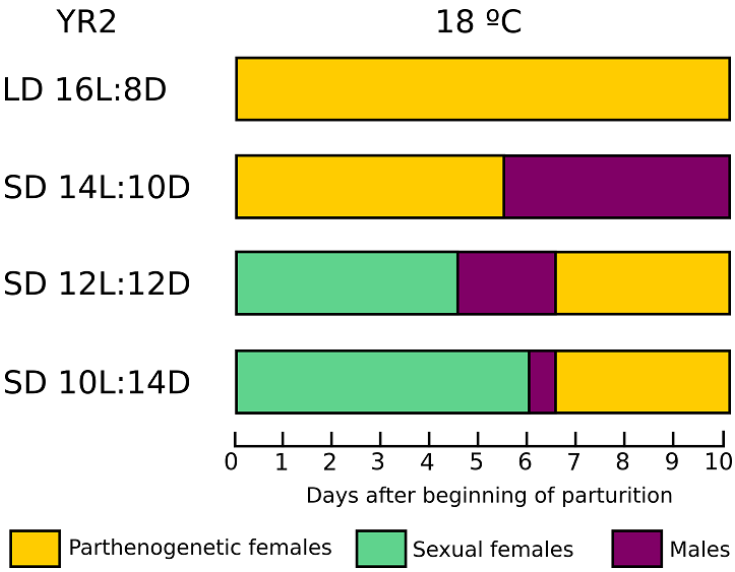


Figure 1.3: Graphical representation of typical birth sequences in *A. pisum* (YR2 strain) exposed to seasonal conditions ranging from summer to winter photoperiods. When photoperiod begins to shorten (14L:10D), males are produced. When conditions are closer to winter-like (12L:12D) sexual females appear. SD, Short Day; LD, Long Day; YR2, "York Red 2" strain of *A. pisum*; 16L:8D, 16 hours of light and 8 of darkness. Adapted from Cortés (2010).

there is a common general pattern (Figure 1.3): under long summer photoperiods parthenogenetic females are born.

The aphid photoperiodic response varies with latitude . Populations found at higher latitudes are exposed to more severe seasonal changes in environmental conditions. *A. pisum* populations living closer to the pole switch to sexual reproduction already at longer days as compared to populations living closer to the equator (Smith & MacKay, 1990). Some populations inhabiting particularly mild,

more constant environments can present an extreme phenotype: they don't respond to shortening of the photoperiod. These populations are termed "anholocyclic", as they don't show a full cycle. They can be considered as life cycle natural mutants or parthenogenetic mutants. Anholocyclic strains reproduce parthenogenetically every season, never producing sexual morphs. Aphids presenting the full cycle, producing sexual morphs and eggs are known as "holocyclic". Despite the fact that anholocyclic lineages don't give birth to a sexual generation, it has been seen that some males might still be occasionally produced, which can mate with the sexual females from the holocyclic populations, thus transmiting the anholocyclic mutation(s).

This pattern of sexual morphs emergence allows for genetic material exchange between different populations. Males, which are frequently winged, might migrate to other plants where oviparae from other populations are being born, thus reducing the probability of inbreeding. When aphids reared under laboratory conditions are exposed to shorter photoperiods, the progeny varies accordingly. Our laboratory stocks includes colonies of the *A. pisum* holocyclic strain named YR2 ("York Red 2" collected originally in York, UK). YR2 has been the reference holocyclic strain for years in our laboratory. The stock is being kept under long days (LD) conditions. The LD illumination regime consists of 16 hours of light and 8 of dark-

ness (16L:8D), which doesn't induce a sexual response in YR2, thus reproducing parthenogenetically. When photoperiod is shortened by two hours (short day "SD" conditions: 14L:10D), YR2 produces males. Finally, when photoperiod is more drastically reduced to shorter conditions (SD: 12L:12D or 10L:14D) YR2 produces different proportions of sexual females. Hereafter we will refer to short photoperiods (14L:10D and 10L:14D) as "SD_{14}" and "SD_{10}" respectively. Although their life cycle has been thoroughly studied, the molecular mechanism governing the photoperiodic response ending in the production of diapausing eggs is largely unknown, not only in aphids, but in insects in general. *A. pisum* is a better model to study photoperiodic responses than the model insect *Drosophila melanogaster*, as the latter displays a very weak seasonal response (Koštál, 2011). However, the availability of genetic resources and tools is much more limited in aphids. The International Aphid Genomics Consortium (IAGC) made available the sequence of the pea aphid genome (The International Aphid Genomics Consortium, 2010), thus promoting advances in studies of the molecular processes and genes involved in numerous aspects of aphid biology. The genome has around 540 Mbp arranged in four chromosomes. It contains 37000 predicted transcripts (Brisson et al., 2016), a high number compared to other insects. In the Aphidbase database[1]

[1]http://www.aphidbase.com/aphidbase

created for the aphid genome project, *Acyrthosiphon pisum* gene IDs begin with acronym "ACYPI" followed by 5-6 numbers (e.g. ACYPI005768 corresponds to a gene enconding the photosensitive protein Cryptochrome 1). Some relevant features of the pea aphid genome include the duplication of more than 2000 gene families and the loss of some conserved genes when compared to other sequenced organisms. Examples are the immune deficiency (IMD) pathway and the urea cycle (The International Aphid Genomics Consortium, 2010).

Apart from parthenogenetic and sexual morphs, aphids present other discrete variations in morphology which can emerge from the same genetic background. That is to say, individuals can vary drastically in their phenotypes (morphologically or physiologically). This is known as polyphenism and aphids make a valuable object of study of this phenomenon. Some environmental factors such as overpopulation, poor food quality or presence of predators can cause the mother to start producing winged progeny to allow exploration of different feeding sources or escape threats (Figure 1.4). In some aphid species, males can also be both winged or not. Wingless individuals are known as "apterous". Body colour, shape, antennal segmentation and other traits can also vary between individuals (Salazar et al., 2015).

Figure 1.4: Aphids attacked by *Aphidoletes aphidimyza*. The predatory midge *Aphidoletes aphidimyza* (center) is one of the various species praying on pea aphids. The presence of predators is one of the many stimuli that can trigger a response in the phenotype of the next generation to develop wings. Photo by Whitney Cranshaw, Colorado State University / CC BY 3.0.

1.2 Biological rhythms

Biological rhythms are cyclic changes in the physiology and behavior of organisms that appeared early in evolution as adaptions to periodical changes in environmental conditions. Among others, these changes include variations in temperature, light exposure, humidity, nutrient availability and presence of predators. Two biological rhythms are most prominent: daily and seasonal rhythms.

Certain daily rhythms are referred to as circadian rhythms and adapt organisms to the day-night cycles caused by Earth's rotation on its axis. The term "circadian" derives from latin, as *circa* means "close to" or "around" and *dies* means "day", as these cycles have a periodicity of about 24 h. Some examples are the sleep/wake cycle in humans and other animals or flower opening in some plants (van Doorn, 2003). Circadian rhythms result from internal processes driven by an endogenous circadian clock (see section 1.3.1), but are entrained by external cues, such as light. Light resets the clock, thus synchronizing physiology with the environment. These cues are known as *zeitgebers* or "ZT", from the german word meaning "time giver".

Seasonal rhythms represent adaptations to environmental changes generated by Earth's revolution around the sun and are most conspicuous in organisms living in temperate regions of the planet as

environmental differences along the year are more marked outside the tropics (Figure 1.5). Organisms have adopted different strategies to cope with adverse conditions. When certain environmental cues announce the beginning of a new season organisms modify their physiology or behave accordingly. Migration, floration, reproduction or diapause are well known seasonal behaviours. The main source of information for seasonal change is day length (photoperiod) , as it is highly reliable and cyclical, unlike temperature, which can vary from one year to the next and whose variations depend on many environmental factors (Figure 1.6).

When winter is close, days become shorter and most organisms can detect the difference and prepare for the upcoming new conditions. Photoperiodism is defined as the ability shown by many species to use day length (photoperiod) to respond to seasonal changes (reviewed in Nelson et al, 2010). It is frequent, however, for organisms to measure the length of the night rather than that of the day. Photoperiodism was first described in plants (see section 1.3) and the first animal to be described as photoperiodic was the insect *Aphis forbesi*, an aphid. Since then, aphids have been one of the most studied animal groups in photoperiodic response. The pea aphid *Acyrthosiphon pisum* and the vetch aphid *Megoura viciae*, as most aphids, switch to sexual reproduction when days shorten (see section 1.1 for details).

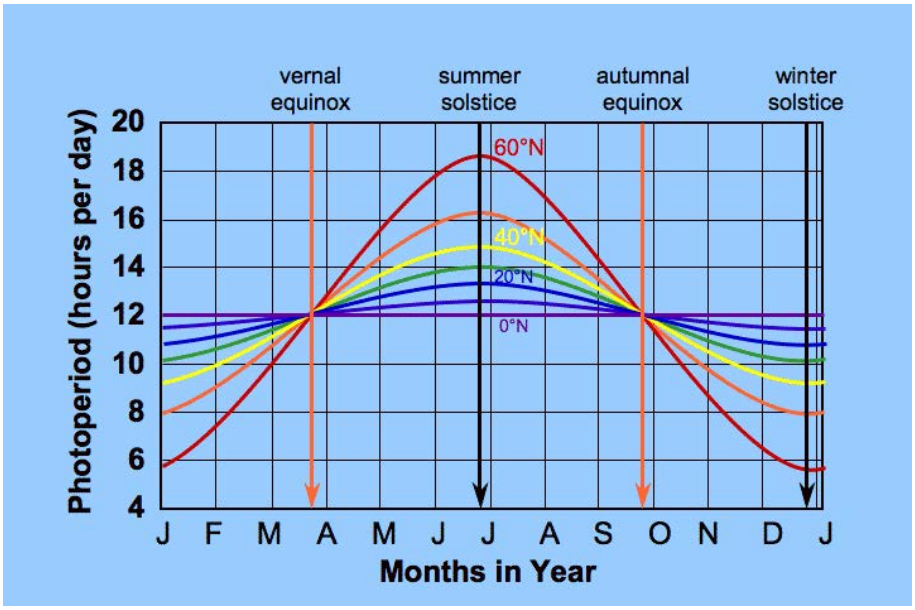


Figure 1.5: Photoperiod along the year at different latitudes. Each latitude is represented by a colour line. At 0ºN (the equator) there is no difference in day-length during the year. At higher latitudes, there are less tan 6 hours of light in January, while more than 18 hours of light in July (at 60ºN). Taken from Koning (1994)

Contrary to the circadian clock, the photoperiodic response is not an endogenous process. The circadian clock keeps ticking with a 24h period even under constant darkness. The seasonal photoperiodic response, however, is triggered (and only takes place) when the corresponding external stimulus (day or night length variation) is sensed. It is worth mentioning, though, that endogenous annual cycles do exist. One of the few reliable examples of such circannual rhythms in insects is found in the pupation rhythm of the carpet beetle, *Anthrenus verbasci* (Miyazaki et al., 2005).

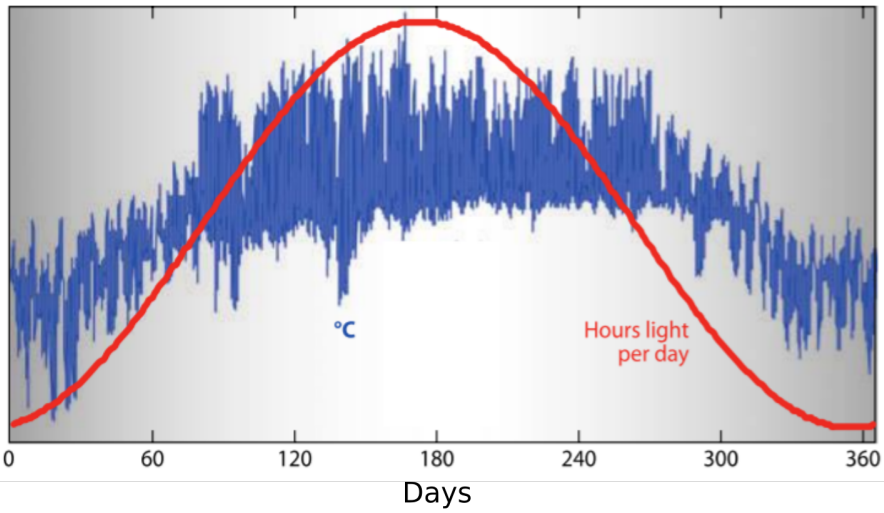


Figure 1.6: The annual fluctuation in temperature and day length. Temperature (in blue) shows important and unpredictable variations even when measured in the same location along the year. Day length, on the other hand (represented with a red line), increases or decreases steadily and reliably year after year as it depends of astronomical events rather than local variations. Temperature measured every 2 hs inside a pitcher plant for one year at 30ºN (Florida, USA). Adapted from Bradshaw and Holzapfel (2010).

Unlike the circadian clock, the molecular mechanism involved in photoperiodic response in insects is still poorly understood. Different models have been proposed: some postulated the presence of two different mechanisms for circadian rhythms and photoperiodism while others imply a coordination of both systems or even the existence of only one governing both phenomena (Vaz Nunes & Saunders, 1999). In all cases, the molecular mechanism underlying photoperiodism in insects is still mostly unknown. How is light

detected and what changes occur in the insect brain during induction of a response is, for the most part, still a mystery.

The photoperiodic response in vertebrates is better understood. Mammals and birds, have certain differences, but both of them make use of the circadian clock in their respective photoperiodic responses (Follet & Sharp, 1969; Goldman, 2001) (Figure 1.7).

The circadian clock of mammals is linked to the photoperiodic response through the pineal gland and the hormone melatonin (Goldman, 2001; Yasuo & Yoshimura, 2009; Ikegami & Yoshimura, 2012) (see Figure 1.7). Melatonin secretion occurs at night, its synthesis is circadian and depends on light availability.

In mammals, light is received through the eyes in the retinal photoreceptors. The light signal reaches the suprachiasmatic nucleus (SCN), which houses the mammal circadian clock and induces the pineal gland via sympathetic innervation to produce melatonin. Because it is synthesized during the night, melatonin secretion varies with the length of the night. Melatonin binds to its receptors in the pars tuberalis ("PT", a part of the pituitary gland) which, in turn, releases thyroid stimulating hormone (TSH). TSH acts on the ependymal cells (EC) in the mediobasal hypothalamus (MBH) to induce the expression of *Dio2* and *Dio3* genes. Enzymes DIO2 and DIO3 locally activate and deactivate thyroid hormone (TH) respectively to fine-tune its levels in the MBH, as it triggers

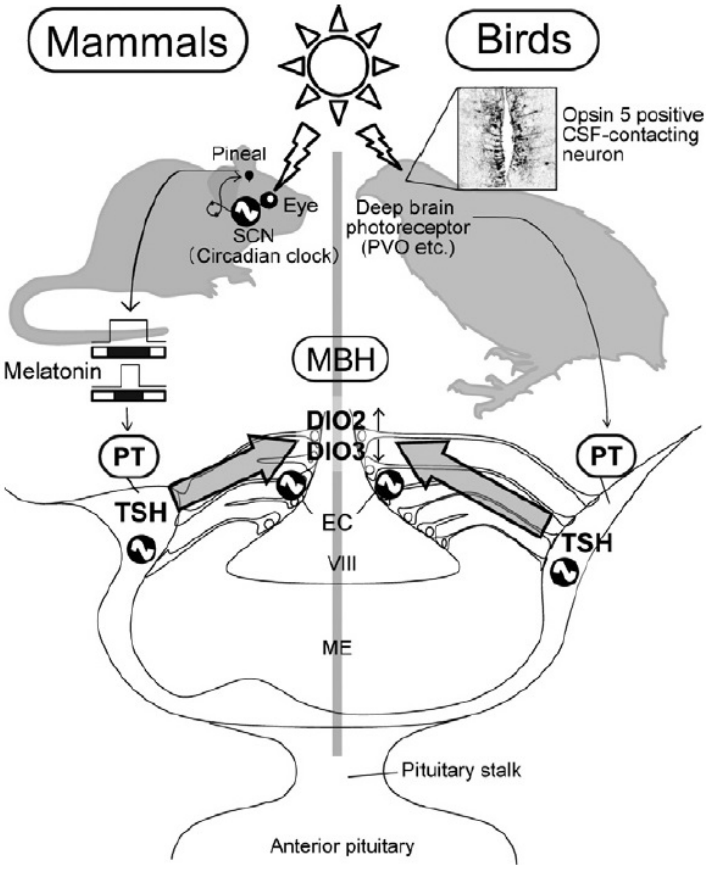


Figure 1.7: Photoperiodic signal transduction cascade in vertebrate controlling reproductive status. In mammals, light is detected through the retinal photoreceptors in the eye. The signal is sent to the suprachiasmatic nucleus (SCN) and pineal gland, where melatonin is produced depending on the photoperiod (higher levels at night). Melatonin levels affect thyroid stimulating hormone (TSH) in the pars tuberalis (PT) which triggers the release of hormonal factors in the mediobasal hypothalamus (MBH). In birds, the deep brain photoreceptor Opsin 5 is expressed in neurons contacting the cerebrospinal fluid (CSF). PVO: paraventricular organ; EC: ependymal cells in the MBH; DIO2 and DIO3 enzymes fine-tune the thyroid hormone (TH) concentration in the MBH. Taken from Ikegami and Yoshimura (2012)

gonadotrophin-releasing hormone (GnRH). GnRH is released to the blood and trigger diverse responses depending on the photoperiod that triggered the cascade of events.

The system is different in birds, where the signal to the PT is not mediated by the eyes eliciting melatonin release. Rather, light is sensed directly in the paraventricular organ (PVO) with deep brain photoreceptors (e.g., protein Opsin 5). Contrarily to mammals, the eyes, the SCN and melatonin are not relevant in bird photoperiodic response (Ikegami & Yoshimura, 2012).

Information available in vertebrates provides us with candidate genes, pathways and molecules to search in other organisms, such as insects.

1.3 Molecular mechanisms

1.3.1 Circadian clock

Both circadian rhythms and photoperiodic responses are driven by molecular mechanisms that receive external information and trigger the appropriate responses that eventually result in a morphological, physiological and/or ethological strategy according to the time of the day or season. Circadian rhythms where first described by De Mairan (1729) in the plant *Mimosa pudica*, whose leaves open and close rhythmically even in the absence of light. It was not

until 1918 that circadian activity was registered in insects: Roubaud measured a daily rhythm in flight activity of the mosquito *Anopheles maculipennis* (Roubaud, 1918; reviewed by Rivas et al., 2016). The first mutant flies (*Drosophila melanogaster*) with circadian rhythm disorders were found by Konopka and Benzer (1971). Those mutants were used to identify the first gene involved in time keeping (Bargiello et al., 1984; Reddy et al., 1984): *period* (*per* for short), which produces the protein PER. The next gene found was termed *timeless* (*tim*) (Sehgal et al., 1994), and the interaction between the TIM protein and PER forms a feedback loop that takes around 24h. These were the first of many genes known to be a part of the molecular clock in *D. melanogaster* (Figure 1.8) and many of them were later found to be conserved in mice and humans.

The first gene found in mammals, however, was *Clock* (*Clk*). It was first identified in mice (Vitaterna et al., 1994; King et al., 1997a; King et al., 1997b) and then found to be conserved in insects (Allada et al., 1998). These were the first of around a dozen genes known to be involved in insect circadian clocks. *Clk*, *per*, *tim* and a fourth gene, *cycle* (*cyc*) participate in two interlocked feedback loops that form the core of the clock. The products of these genes interact with each other and with molecules outside the clock. Three main elements were proposed to be required in an endogenous clock mechanism: a central oscillator (the core), an input pathway and

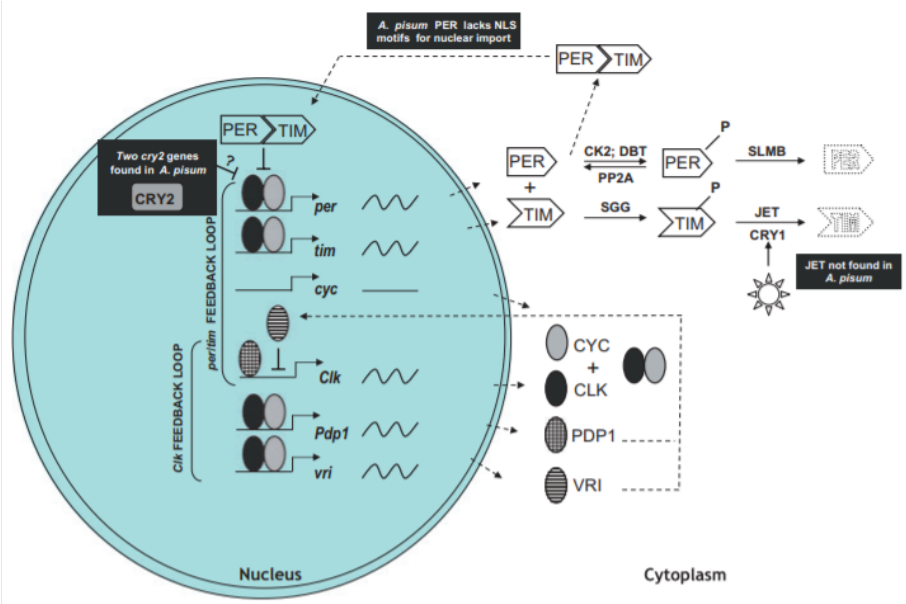


Figure 1.8: The circadian clock in *D. melanogaster*. The clock consists of two feedback loops (Cyran et al., 2003). The genes involved are represented with crooked arrows. Differences with *A. pisum* are shown in black boxes. Core proteins in the feedback loops are represented by different shapes labelled in capital letters. Proteins targeted for degradation have dotted outlines. CRY2 is absent in *D. melanogaster* but duplicated in the pea aphid. The putative role as repressor of CLK/CYC dimers (Yuan et al., 2007) is suggested by a question mark. CLK/CYC dimers and PDP1 attached to crooked arrows indicate their roles as transcription activators of corresponding genes. Lines ending in bars indicate negative regulation. Wavy lines indicate rhythmic transcription in *D. melanogaster*. Taken from Cortés (2010).

the output pathway (Figure 1.9).

Central oscillator

The core or central oscillator, as described above, consists of a series of biochemical reactions that cycle autonomously in periods of (around) 24h. The oscillator receives information through the input pathways, which provide data about the environment, so as to entrain (or synchronize) the central oscillator with the environment. The main and most relevant input pathway for circadian clock entrainment is sunlight, but some organisms rely on others. The cyclical information originating in the oscillator is transmitted to the organism through the output pathways, which regulate rhythmic physiology and behaviour in accordance to the time of the day (see Figure 1.9).

The clock cycles internally with the same rhythm over a wide range of temperatures, which means that it compensates for differences in this variable (Bodenstein et al., 2012). Some circadian systems include secondary loops that modulate amplitude, precision, and stability (Gardner et al., 2006).

The insect circadian clock has been intensively studied in *D. melanogaster*. The 2017 Nobel prize in Physiology or Medicine was awarded to Jeffrey C. Hall, Michael Rosbash and Michael W.

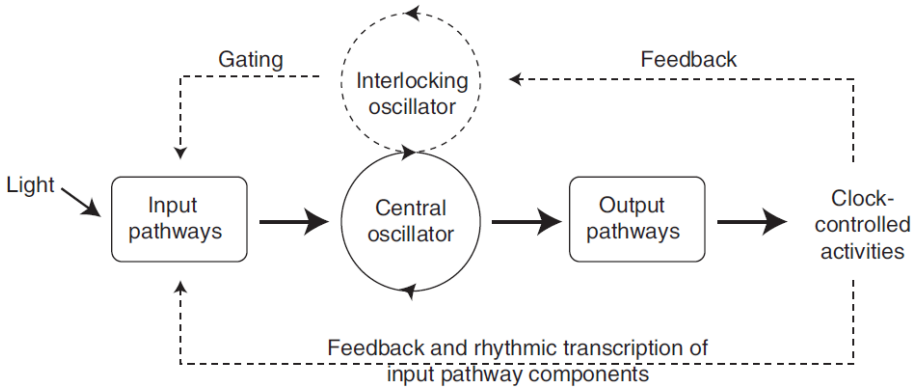


Figure 1.9: The main elements in a circadian clock. The input pathway detects light to entrain the central oscillator. The central oscillator or pacemaker consists of interlocked feedback loops that cycle every 24 hours. The output pathway transmits the information to the rest of the body. Some organisms have secondary interlocking oscillators to fine-tune the central pacemaker. All these elements interact reciprocally, modifying their activity to different extents. Modified from Kuhlman et al. (2017) and Gardner et al., (2006).

Young for their discoveries in this field[2]. As already mentioned, the circadian clock consists of two main interlocked feedback loops. Some of the elements in the loops are activators (positive elements) while others are repressors (negative elements). Two proteins, Clock (CLK) and Cycle (CYC), are positive elements, as they bind forming an heterodimer CLK:CYC that activates expression of other genes. CLK:CYC binds to an enhacer box (E-box), a short sequence upstream of other clock genes like *timeless*, *period*, *Pdp1*, *vrille* and other clock-controlled genes. Proteins Period (PER) and Timeless (TIM), on the other hand, are negative elements. They repress,

[2]https://www.nobelprize.org/prizes/medicine/2017/press-release/

among others, their own gene expression when forming the heterodimer PER:TIM, as PER ineracts with CLK:CYC interrupting CLK:CYC-mediated activation. In a second loop, *Pdp1* (PDP1) and *vrille* (VRI) control *clock* expression. PDP1 activates *clock* expression while VRI represses it.

Other elements of the clock include several kinases and phosphatases: double-time (DBT), shaggy (SGG), protein phosphatase 2a (PP2A), supernumerary limbs (SLMB) and Casein kinase 2 (CK2) regulate degradation, cellular localization and stability of PER and TIM proteins, which is essential for resetting the clock everyday (see below)

The Output pathway

The output pathway of the insect clock has not been completely elucidated. The output pathway in vertebrates is, again, better understood. In vertebrates, which have been more thoroughly studied, melatonin plays, in many cases, a relevant role as one of the output elements of the clock (Klein, 2006). It is an indole amine synthesized from serotonin (5-hydroxytryptamine or 5-HT) via an arylalkylamine N-acetyltransferase (AANAT). AANAT has been referred to as "The timezyme" (Klein, 2006), given that it is the rate limiting enzyme in melatonin synthesis. In vertebrates, melatonin levels increase at night, thus representing a chemical indicator of

internal time. Although the role of melatonin in insect circadian clocks is unclear, there is evidence linking this hormone with insect time measurement. Several insects show increased melatonin levels at night, its precursors and related enzymes are expressed in anatomical regions related to photosensitivity and/or circadian pacemakers (Vivien-Roels & Pévet, 1993). In vertebrates, apart from its role in circadian clocks, melatonin is also involved in the seasonal response (section 1.3.1)

Independently of the putative role of melatonin, there is one group of neuropeptides whose role in the output pathway or circadian clocks in insects is confirmed: the Pigment-Dispersing Factor (PDFs) (Renn et al., 1999; Shafer & Yao, 2014). PDF and PDF receptors are expressed in *D. melanogaster* neurons expressing clock genes (Helfrich-Förster, 2004; Shafer & Yao, 2014). The *pdf* gene mutants show different alterations like absence of the anticipatory morning peak of activity or arrhythmicity (Shafer & Yao, 2014). It is, however, surprising that *pdf* gene has not been found in aphids (Le Trionnaire, et al., 2009; 2012; Barberà, 2017).

Also related to the output pathway, *takeout* gene (*to*) is proposed to be a clock regulated output gene only found in insects (So et al., 2000). This gene is not only under clock control, but it is also overexpressed under starvation conditions. *to* seems to provide temporal and food status information to metabolism and activities

that depend on it (Sarov-Blat et al., 2000).

The Juvenile Hormone (JH), a very relevant molecule in insect development, might also be part of the circadian output (Bloch et al., 2013). JH is rhythmically synthesized in bees and crickets (Elekonich et al., 2001; Zera, 2016). JH hormone is also involved in the aphid photoperiodic response, as it promotes the production of asexual, parthenogenetic morphs (Hardie, 1981; Denlinger et al., 2001)(see below).

The Input pathway: cryptochromes and opsins

Regarding light sensitivity used to entrain the molecular clock, it is particularly relevant to mention the role of the cryptochrome (CRY1) protein, a blue-light sensitive photoreceptor (see section 3.4). In *D. melanogaster*, CRY1 modifies TIM when exposed to light, usually at dawn. Modified TIM binds to Jetlag protein (JET), a ubiquitin ligase which targets TIM for degradation (Lin et al., 2001; Koh, 2006, Peschel et al., 2009). Degradation of TIM resets the clock, thus entraining the system. JET also targets CRY1 for degradation (Peschel et al., 2009).

The Cryptochrome protein described above for *D. melanogaster* is not the only one present in insects. Other insects have an additional Cryptochrome homologue (CRY2), which is closely related

to the Cryptochrome found in mammals (thus often referred to as "mCRY", while CRY1, originally found in *D. melanogaster* is also referred to as "dCRY"). As seen in Figure 1.10, insects can show different combinations of Cryptochromes (Yuan et al., 2007): they can carry only one cryptochrome or both. *D. melanogaster* only has CRY1, but beetles and bees (*Tribolium castaneum* and *Apis mellifera*) only have CRY2. Monarch butterflies (*Danaus plexippus*) have both, CRY1 and CRY2. CRY2 doesn't have photosensitive properties, but acts as a transcriptional repressor within the clock (Yuan et al., 2007) along with PER. *A. pisum* not only has both cryptochromes, but also CRY2 is duplicated. This makes *A. pisum* an interesting subject of study on the molecular mechanisms of circadian clocks. See section 3.4 for further information on cryptochromes and their role in circadian clocks.

Opsins, like cryptochromes, are photosensitive proteins. They are vitamin-A derived proteins used as visual pigments in animals, and structurally similar proteins are used throughout the tree of life (Terakita, 2005; Terakita et al., 2011). These proteins are also used for image forming and are usually expressed in the eyes. Color images are formed by combining absorption spectra from two or three different opsins in dichromatic and trichromatic organisms respectively.

In addition, evidence from different insect species, including

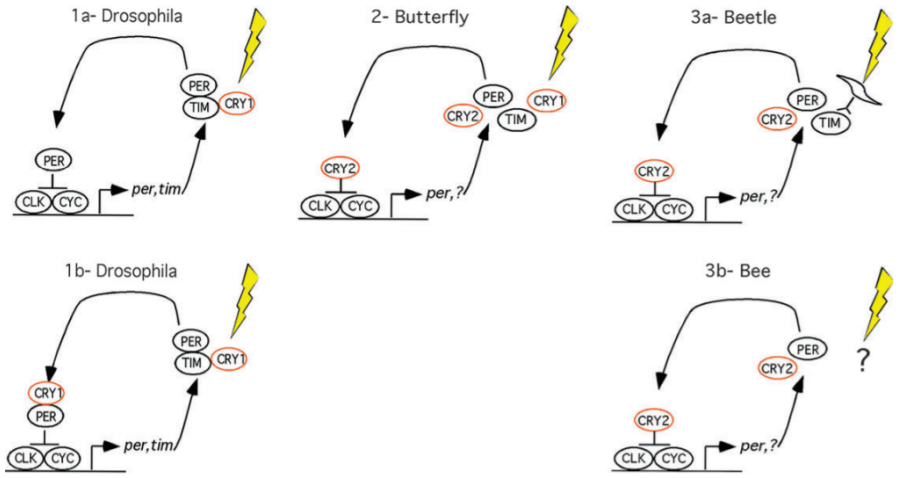


Figure 1.10: Insect clock models based on cryptochromes (*cry*). There are three different types of insect clock models considering the different combinations of *cry1* and *cry2* genes. The first model (1a and 1b) is present in *Drosophila*. In this model only *cry1* is present, and it could adopt two roles: as photosensor (1a) and as central clock component (1b). The second type includes both, *cry1* and *cry2*, and is present in butterflies and aphids. They can perform as photoreceptor and as central clock component respectively. The third type of model involves *cry2* only. It is present in beetles and bees (3a and 3b) and might be involved in both, photoreception and transcriptional repression. Taken from Yuan et al. (2007).

Drosophila, suggests that apart from their involvement in vision-related tasks opsins expressed in the eyes also participate in circadian photoentrainment (Saint-Charles et al., 2016; Helfrich-Förster). As in vertebrates, the opsin family in insects also includes genes expressed not in eyes but in the brain, where they would perform other non-visual light sensing-related tasks yet to be fully elucidated. For instance, a putative role in circadian entrainment attributed to *Drosophila* Rh7 (Ni et al., 2017) was later seriously questioned (Senthilan et al., 2019). A possible role in photoperiodism of some of these opsins outside the eyes has guided some investigations in aphids (Gao, von Schantz, Foster, and Hardie) since a role for the eyes was discarded by early investigations (Lees).

Each cryptochrome and opsin is more sensitive to a specific range of light wavelengths. Lees (1973; 1981) performed a series of experiments that better determined the wavelength spectrum that triggers a photoperiodic response the aphid *Megoura viciae*. He exposed aphids to short day (long night) conditions and interrupted the dark period with light pulses at different times of the night and with lights of different wavelengths. When the aphid perceived the light pulses at certain moments of the night, it was sensed as a long in stead of short day, thus interrupting the diapause induction. Lees (1973; 1981) found that in *M. viciae* there are two moments of the night at which the light pulse is perceived as a long day: early in

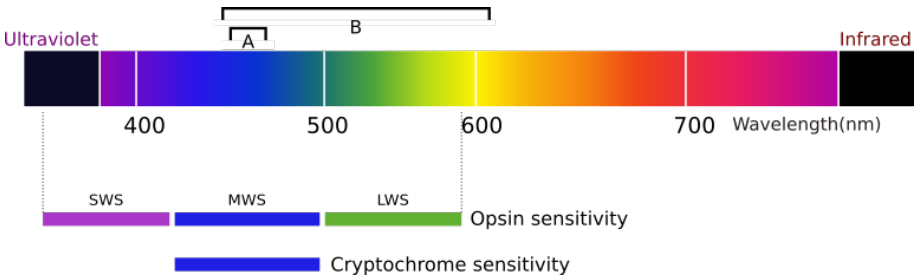


Figure 1.11: Approximate light wavelength sensitivity of insects opsin genes and cryptochromes. A: light wavelength range that elicits a long day response in the aphid *Megoura viciae* when exposed at the early night (Lees, 1973). B: light wavelength range that elicits a long day response in the aphid *M. viciae* when exposed at the late night. SWS: short wavelength sensitive; MWS: medium wavelength sensitive; LWS: long wavelength sensitive

the night and also one point late in the night. The first moment is sensitive to blue light (450-470 nm) (Figure 1.11 while the second moment is sensitive to a broader spectrum, from blue up to 600 nm (close to red). Thus different photoperiodic photoreceptors are probably being used at different moments of the night (reviewed by Saunders, 2012). Cryptochromes are blue-sensitive while different opsin genes' sensitivity ranges from short (ultra-violet) to medium (blue) and long (green-yellow) wavelengths.

1.3.2 Photoperiodism in aphids

Although the molecular bases behind the circadian clock have been studied and characterized to a certain extent (see previous section), the mechanism underlying the seasonal response in insects

is less understood. A molecular mechanism such as that proposed for the circadian clock is not available. There are, however, several theoretical models (Vaz Nunes et al., 1999; Tauber et al., 2001; Denlinger et al., 2001; Danks, 2003). Once again, it was research in plants the starting point for understanding a biological rhythm and laying the bases for the aforementioned models. Turnois and Klebs proposed in the early 1910's that the duration of the light phase during the day is a key factor in plant development (Klebs, 1910, 1913; Tournois, 1912, 1914; Thomas & Vince-Prue, 1996). In 1920 Garner and Allard (Garner & Allard, 1920) discovered that tobacco plants (*Nicotiana tabacum*), flowered at the same time guided by the length of the day. They introduce the terms "photoperiod" and "photoperiodism". Shortly after, Marcovitch (Marcovitch, 1923, 1924) experimentally determined for the first time the photoperiodic induction of an animal, the aphid *Aphis forbesi*. Similar to *A. pisum*, when exposed to short days *A. forbesi* responds by producing sexual morphs (see Figure 1.1).

By analogy with the circadian clock, the proposed photoperiodic clock model (Denlinger et al., 2001) has some similarities and differences when compared to the circadian clock model (Figure 1.12): there should be an input pathway, a core mechanism and output pathways. In general terms, photoperiodic organisms require (Denlinger et al., 2001):

1. A photoreceptor mechanism to sense the light
2. A clock mechanism to measure day or night length
3. A photoperiodic counter to keep track of the number of short (or long) days passed
4. A neuroendocrine effector mechanism to coordinate the response in the organism

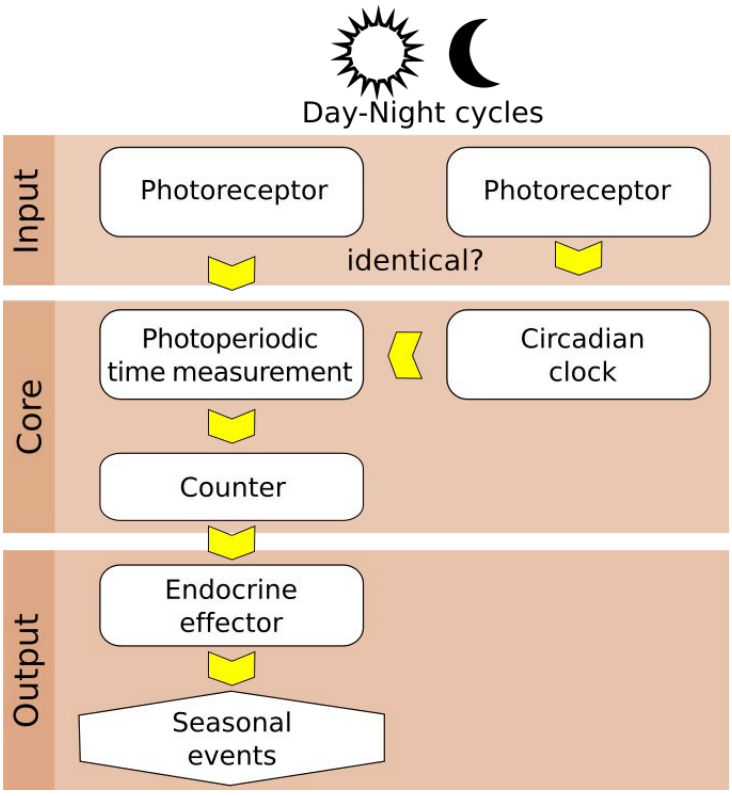


Figure 1.12: Elements involved in a putative photoperiodic system. Light stimulus is detected by photoreceptors, which might be shared by the photoperiodic response and the circadian clock. The core of the photoperiodic clock is responsible of measuring the length of the day and of keeping track of the accumulated long (or short) days passed. When a given number of days with certain photoperiod length signals an upcoming season change, endocrine effectors are released to induce the proper response in the organism. Taken from Hoffmann (2014)

Photoreceptor mechanism

The first studies aimed at determining the localization of the photoperiodic photoreceptors in aphids concluded that photoperiod was sensed through the head of mother (Figure 1.13) (Lees, 1964). This stimulus is transformed into a physiological message that modifies the development of the progeny. The experiment discarded the possibility that photoperiod is directly detected by embryos through their mother's abdomen. Ruling out this possibility was particularly relevant, as developing embryos are located in the abdomen, and light might have been going through the parthenogenetic mother's cuticle directly to the embryos' photoperiodic receptor. That, however, is not the case.

The experiment performed by Lees (1964) tested the effect of illuminating specific parts of the anatomy of *M. viciae* aphids reared under SD conditions with extra periods of light. Thus, the illuminated parts of the body experienced LD (long day) photoperiods, while the rest of the body perceived SD (short day) conditions. As seen in Figure 1.13, 100% of aphids receiving LD conditions in the dorsal region of the head produced parthenogenetic females, the expected progeny under this photoperiod. On the contrary, aphids receiving extra illumination (LD condition) in the abdomen failed to produce parthenogenetic females. Although targeted illumination

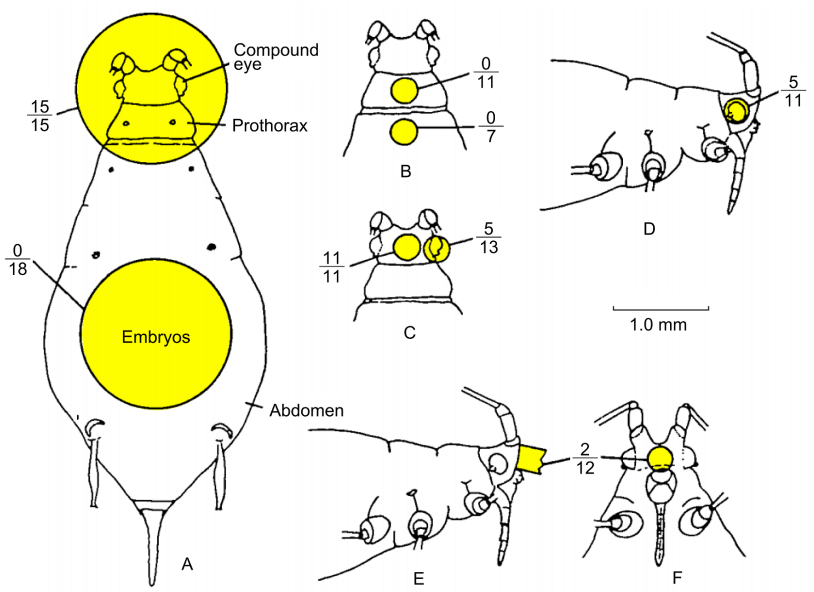


Figure 1.13: Results of localized illumination experiments on *Megoura viciae* body regions. In yellow circles, regions receiving extra illumination, locally extending the photoperiod from Short Day (SD) to Long Day (LD). Fractions indicate the number of aphids reverting to parthenogenetic reproduction over the total of individuals assayed as a response to LD conditions. Modified from Lees (1964).

in the eyes does elicit a certain response to LD photoperiods, it is less effective, and the authors proposed that the positive response could be caused by light scattering while illuminating the eye.

A more specific localization of photoperiodic photoreceptors was achieved in a classical experiment using microlesions apart from targeted illumination of head regions of *M. viciae* (Steel, 1978) (Figure 1.14).

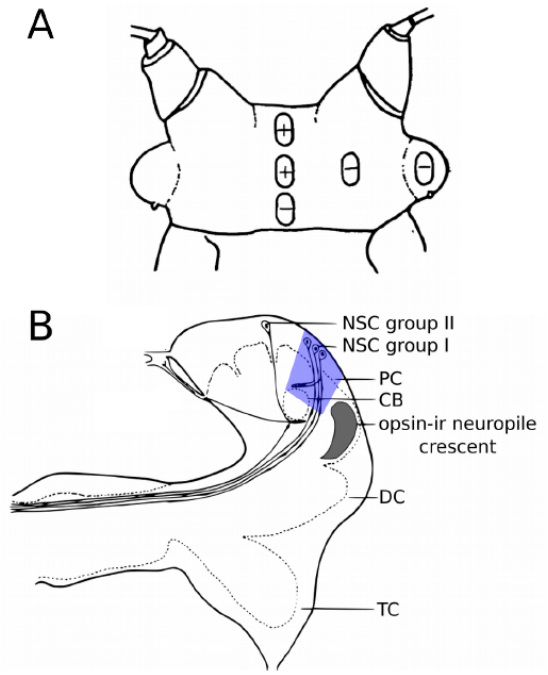


Figure 1.14: Localisation of the putative photoperiodic photoreceptors in *Megoura viciae*. **A)** Dorsal view of the aphid head showing light sensitive regions. Plus symbols: regions exposed to long photoperiod that reverted oviparae production. Negative symbols: Regions that didn't revert oviparae production. Adapted from Steel (1978). **B)** Near sagital section of the aphid brain showing the location of the opsin immunoreactive region (opsin-ir) (Gao et al., 1999). The essential pars intercerebralis region for the photoperiodic response is shown in blue. If the region in blue is affected with a microlesion, the photoperiodic mechanism is disrupted (Steel and Lees, 1977). PC, protocerebrum; CB, central body; DC, deutocerebrum; TC, tritocerebrum; NSC, Neurosecretory Cells identified by Steel (1977b), see section 1.3.1. Adapted from Gao et al. (1999).

The aphid brain

The brain is the main coordination site of photoperiodic and circadian activity in the aphid. It is thus relevant to describe briefly the main components of the aphid brain (Figure 1.15). The central nervous system (CNS) is formed by three main elements, including the brain, the subesophageal ganglion (SOG) and the thoracic ganglionic mass (TGM), which extends beyond the head, into the prothorax. Most relevant to this thesis is the brain, which is found in the dorsal region of the head. It is subdivided into the protocerebrum, the deutocerebrum and the tritocerebrum. The protocerebrum includes the optic lobes (which are directly connected to the compound eyes), the central complex, lateral accessory lobes and the mushroom bodies. The central complex includes the central body and the protocerebral bridge (see Figure 1.15). As several genes, proteins and other molecules related to biological rhythms have been detected in other insects' brains, it is relevant to know if this is also the case in aphids.

Regarding the localization of photosensitive proteins in the brain, some studies suggest that proteins belonging to the opsin family can be found in the pars intercerebralis region as revealed by immunoreactivity experiments (Gao et al., 1999) (see Figure 1.14 B). More specifically, in the anteroventral part of the neuropile

(synapse-rich region) in the protocerebrum (Gao et al., 1999), where photosensitive proteins were found. Moreover, opsin involvement in photoperiodism is supported by wavelength sensitivity studies (Lees, 1981).

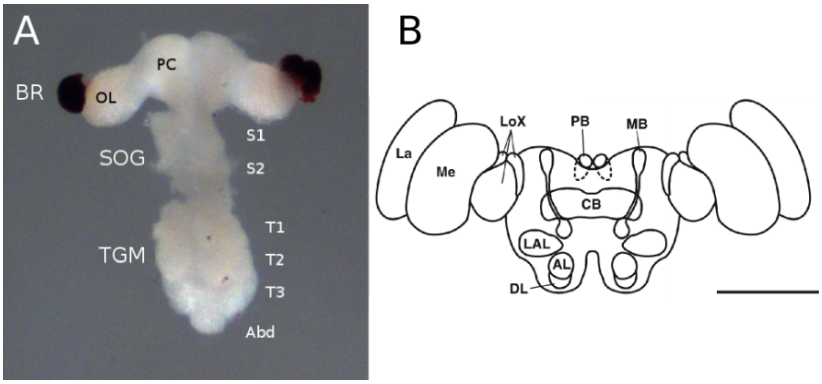


Figure 1.15: *A. pisum* brain. **A)** Disssected aphid Central Nervous System. BR, brain; SOG, the suboesophageal ganglion; TGM, thoracic ganglionic mass. The regions corresponding to the SOG and the TGM are respectively indicated with numbers (S1, S2, T1, T2 and T3). Taken from Barberà (2017). **B)** Anterior view of a representation of the brain and selected neuropils. AL, antennal lobe; CB, central body; DL, dorsal lobe; La, lamina; LAL, lateral accessory lobe; LoX, lobula complex; Me, medulla; MB, mushroom body; PB, protocerebral bridge. Bar = 100 µm. Taken from Kollmann et al. (2010).

Spectral sensitivity studies also support the use of cryptochromes as photoperiodic receptors (Saunders, 2012). See section 3 for detailed information on opsins and cryptochromes and their involvement in aphid photoperiodism. It seems, however that opsins expressed in the compound eyes are not in charge of light sensing

in photoperiodism (Lees, 1964; Steel and Lees, 1977). It could be that other proteins from the opsin family expressed in other tissues are responsible for this task.

Clock mechanism, photoperiodic counter and relation with the circadian clock

It is necessary for the photoperiodic response to measure the length of the day and to keep track of the number the short days or long nights experienced. These functions are performed by the proposed clock and counter of the putative photoperiodic system.

Among the several models proposed (Vaz Nunes et al., 1999; Tauber et al., 2001; Denlinger et al., 2001; Goto, 2012) some have been more thoroughly developed. They can be classified according to the proposed role of the circadian clock in the process: some consider that the photoperiodic clock is dependent of the circadian clock while others believe that both clocks are independent.

The hourglass model (Figure 1.16) considers that the photoperiodic clock is independent of the circadian clock. In this model, some (unknown) substance would accumulate during the day and would be degraded at night, like an hourglass being reset every day. If days are long enough, then the substance levels reach a threshold necessary to trigger a long-day response (see Figure 1.16). Similarly, the hypothetical substance could be accumulated at night

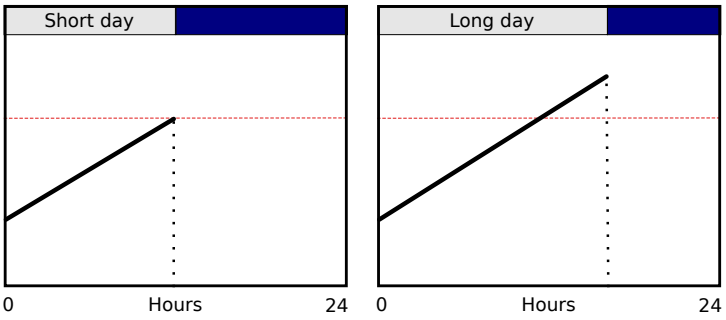


Figure 1.16: Hourglass model. A hypothetical substance accumulates during daytime, but is degraded at night, resetting the hourglass clock. If days are long enough, the substance levels could reach threshold levels (red line) needed to trigger a photoperiodic response.

and degraded at daytime.

The circadian clock-dependent models are classically divided into the external and the internal coincidence models:

1. External coincidence model. Requires the existence of an internal rhythm in sensitivity to light. The organism would be particularly sensitive to light at a given time close to the end of the day. If at that moment it is already night time, then that day is sensed as a short day and photoperiodic response is triggered, provided it is repeated for a certain number of days. On the contrary, if the photosensitive phase meets daylight, then that day is detected as long. In this model light has two roles: it entrains the internal photosensitivity cycle and also serves as stimulus in the photosensitive phase (Figure 1.17).

2. Internal coincidence model. This model implies the existence of (at least) two internal oscillators (C. Pittendrigh, 1966; Saunders, 2009). If their oscillations are synchronized at a critical point, then the photoperiodic response takes place. Both oscillators would be affected differently by light, thus that critical point could eventually be reached. This model requires light to have only one role: entrainment of both oscillators (see Figure 1.17).

In the case of aphids, there is contradictory evidence supporting both the hourglass model and the circadian dependent models. Nowadays, however, another proposal is supported by recent evidence (Beer et al., 2017) and unifies previous contradictory results: the damping circadian oscillator model (Lewis et al., 1987; Saunders, 2002; Goto, 2012; Beer et al., 2017). In general terms, this model makes use of an external coincidence model of the circadian clock, but it consists of an oscillator whose amplitude dampens very quickly in the absence of entrainment (i.e. in the absence of a circadian rhythm). This model predicts that a damping circadian oscillator would behave like an hourglass, thus unifying both proposals. Evidence in aphids shows that some circadian clock genes are affected by photoperiod (Barberà, 2017), thus supporting its involvement in the photoperiodic machinery. Therefore, the circadian clock

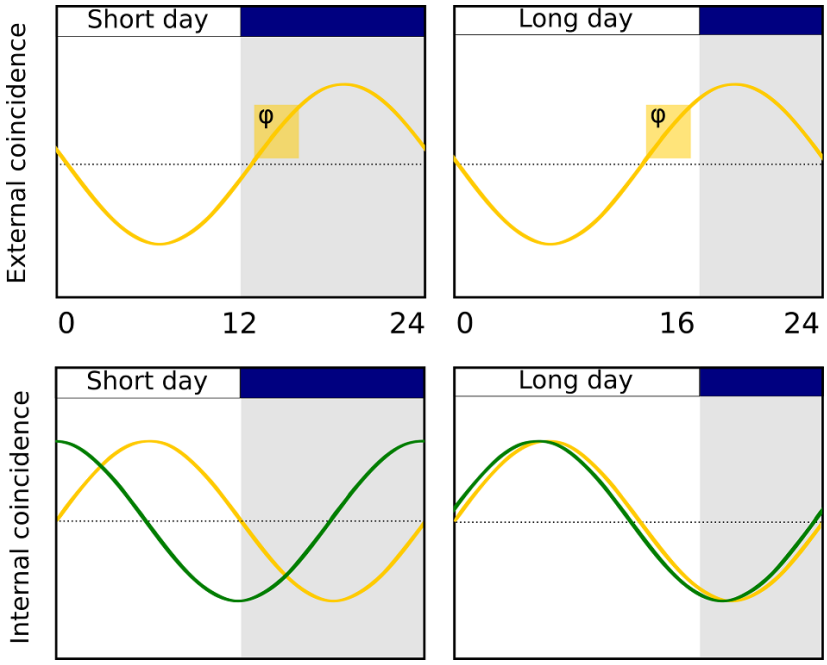


Figure 1.17: **Top**: External coincidence model. When the photosensitive phase of an internal oscillator (φ) is exposed to light, the organism presents a long day response. **Bottom**: Internal coincidence model. Two internal oscillators have to synchronize in order to trigger a response.

provides multiple candidates worth studying in the photoperiodic response.

The output: A neuroendocrine effector mechanism

After short days are detected in the brain and their accumulation reaches a threshold, the message has to be transmitted to generate a response in the ovaries, so as to switch the reproductive phenotype of the developing offspring. This message transmission should involve neuroendocrine effectors and/or neurotransmitters. Early studies in aphids have shown that the neuroendocrine control is mediated by brain neurosecretory cells and the corpora allata (CA), endocrine glands in the brain (Hardie, 1984; Denlinger et al., 2001). Among the possible effectors we can find juvenile hormone (JH), melatonin, prothoracicotropic hormone (PTTH), insuline like peptides (ILPs), dopamine and other small peptides.

Juvenile hormone (JH) is an acyclic sesquiterpenoid that regulates many aspects of insect development. When in excess, it prevents insects from reaching adulthood and sexual maturation. JH is produced in the CA and promotes the production of asexual morphs in aphids (Hardie, 1981; Denlinger et al., 2001). When aphids exposed to SD (which normally induces sexual morphs production) are exposed to JH analogs, they revert to parthenogesis. Evidence also suggests that JH levels are high during long days and

mimic long day effects in aphids (Hardie et al., 1985; Ogawa et al., 2014).

Another hormone related to time measurement is melatonin, which is usually synthesized during the night. In vertebrates, the pineal gland releases melatonin and increases its levels at night. It is known to be crucial for the seasonal response in different animals (Klein, 2006; Walton et al., 2011), as it transduces photoperiodic information (higher levels are attained during short days).

Presence of melatonin in insects is controversial, although it has been detected in some species and also seems to increase at night (Vivien-Roels et al., 1993; Gorbet et al., 2003; Bembenek et al., 2005; Vieira et al., 2005; Yang et al., 2007; Mohamed et al., 2014). It has been proposed that it participates in the photoperiodic response of the lepidopteran *Antheraea pernyi* (Wang Q. et al., 2013b; Mohamed et al., 2014; Wang et al., 2015).

In aphids the evidence is not as strong, yet data suggest a role for melatonin in sexual morph induction (Hardie & Gao, 1997; Denlinger et al., 2001). Thus, it has the opposite effect than JH, which promotes asexual morph production. A recent study by our group (Barberà et al., 2018) has detected differences in melatonin production in the brain of holocyclic and anholocyclic strains of *A. pisum*. The holocyclic strain, which produces sexual morphs, exhibited higher melatonin levels under SD conditions, while the anholocyclic

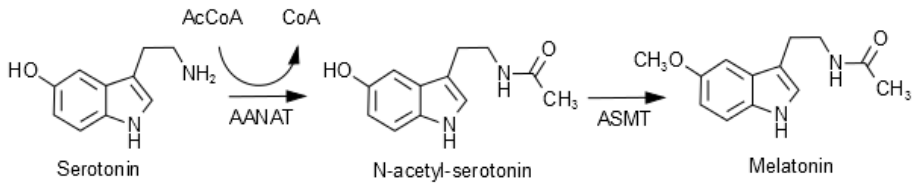


Figure 1.18: Melatonin synthesis from serotonin in vertebrates. AANAT (arylalkylamine N-acetyltransferase) catalyzes the N-acetylation of serotonin to synthetize melatonin, an output of the circadian clock in mammals that is also key for their seasonal response. AANAT is the rate limiting enzyme in the reaction and has been referred to as "the Timezyme" due to its relevance in the process (Klein, 2006).

didn't. In vertebrates, the enzyme AANAT (v-AANAT) is essential in melatonin synthesis from its precursor molecule, serotonin (Figure 1.18). Moreover, insect arylalkylamine N-acetyltransferase (i-AANAT), the enzyme thought to be key in melatonin synthesis in insects (Hiragaki et al., 2015), presents levels of expression compatible with the melatonin levels found (Barberà et al., 2018). Although vertebrate and invertebrate AANATs are not phylogenetically related they are believed to catalyze the same reactions. In *Antheraea pernyi*, for example, melatonin levels seem to be under the control of an i-AANAT (Mohamed et al., 2014), which is in accordance with the possibility of insect AANATs being essential in the production of melatonin.

Serotonin itself has been studied for its effects on circadian rhythms, photosensitivity and seasonality. Serotonin causes phase

shifts in the circadian clock of molluscs and cockroaches (Eskin & Maresh, 1982; Page, 1987). Seasonal response in birds is also affected by daily injections of serotonin during the photosensitive phase of the annual reproductive cycle (Bhatt & Chaturvedi, 1992).

Dopamine (DA) is a neurotransmitter present all along the tree of life that participates in diverse processes in insects (Verlinden, 2018). It has several roles in development. In relation to the photoperiodic response, DA induces a diapause-like state in the lepidopteran *Mamestra brassicae* (Noguchi & Hayakawa, 1997), it blocks diapause termination in *Antheraea pernyi* (Wang et al., 2014) and DA levels are significantly different in diapausing than non-diapausing *Pieris brassicae*, another lepidopteran (Isabel et al., 2001; Tauber et al., 2001). Wang et al. (2015) proposed that the photoperiodic counter could be driven by mutual inhibition between the melatonin and dopamine pathways. All of this cases are in accordance with a decrease of dopamine during photoperiodic response (despite that some photoperiodic responses are triggered by long and others by short days).

Prothoracicotropic hormone (PTTH) plays a key role in the control of insect molting and metamorphosis through the stimulation of ecdysone from the prothoracic glands. PTTH also promotes diapause (which is a frequent response in photoperiodism) (Zhang Q. & Denlinger, 2011; Bloch, et al., 2013), is an output of the

circadian clock (Vafopoulou & Steel, 2014; Mohamed et al., 2014) and is believed to play a role in photoperiodism in aphids (Barberà & Martínez-Torres, 2017).

The insulin signaling pathway, including insulin-like peptides (ILPs) and insulin receptors (IR), has been shown to be involved in diapause in species like the nematode *C. elegans*, *D. melanogaster*, the mosquito *Culex pipiens* and the pea aphid (*A. pisum*) (Apfeld & Kenyon, 1998; Tatar & Yin, 2001; Sim & Denlinger, 2013; Barberà et al., 2019). ILPs receive information from the core clock elements, *per/tim*, and, in turn, affect JH synthesis and FOXO signaling pathways via the IR (Figure 1.19). These pathways affect diapause and development, which is a common response to photoperiodic changes. It has been proposed that insulin signals are likely used as mediators of a wide range of environmental cues, including diapause (Sim & Denlinger, 2013). The role of the various elements involved in the insulin pathway in aphid photoperiodic response is unknown.

Steel and Lees (1977) and Steel (1977; 1978) performed a series experiments to determine the location of aphid neurosecretory cells (NSC) in the aphid brain and identified two relevant and adjacent regions: the *pars lateralis* and the *pars intercerebralis*. Using PAF[3] staining, Steel (1977b) identified five groups of NSCs. The authors

[3]Paraldehyde Fuchsin stains proteins containing disulphide bridges an has traditionally been used for detection of neurosecretory elements in insects (Ewen, 1962)

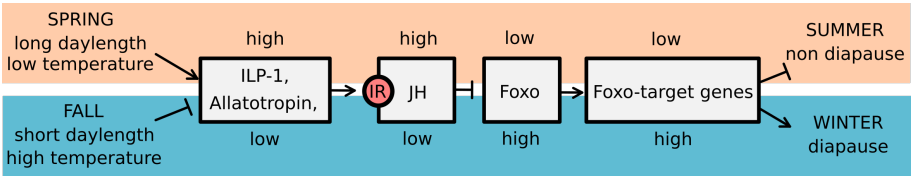


Figure 1.19: Molecular cascade in mosquito diapause. Under short days conditions, neuropeptides allatotropin and insulin-like peptide 1 (ILP-1) levels decrease. Insulin receptors (IR) mediate juvenile hormone synthesis (JH) and, consequently, the Foxo signaling pathway. This cascade of events leads, ultimately, to the overwintering adult diapause in mosquitoes. Adapted from Sim et al., (2013 and 2015).

performed microlesions in the brain, specifically removing particular groups of NSCs. One group (NSC group I) located in the anterior dorsal region known as *pars intercerebralis* was identified as the producer of an effector substance that promotes the production of virginoparae (Figure 1.20). The authors proposed that it might contain the photoperiodic photoreceptor and the clock. Indeed, when they ablated this group of NSCs, aphids reared under SD conditions (that should, then, produce sexual morphs) reverted to the production of parthenogenetic females (virginoparae), thus they named this effector "virginoparin".

Moreover, these ablation experiments also determined that the region in the vicinity, known as *pars lateralis* (see Figure 1.20), is also involved in the process, probably through photoreception or integration of the photoperiodic message.

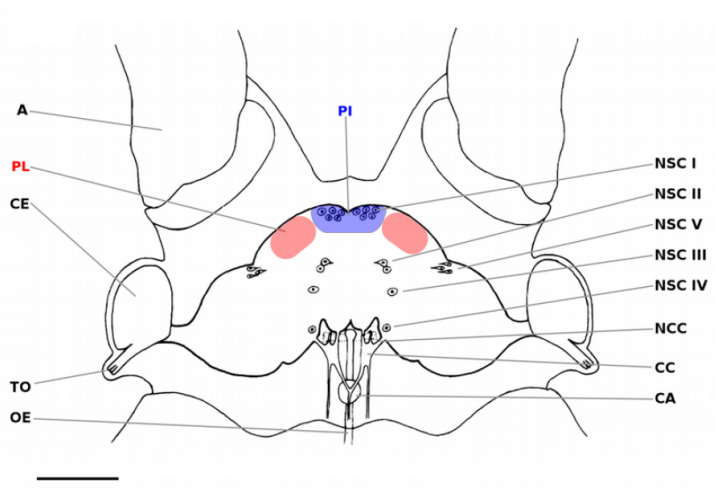


Figure 1.20: Five groups of neurosecretory cells (NSC) in the aphid brain. Dorsal view of the brain in which the position of different NSC (I to V) groups is shown. Destruction of *pars intercerebralis* (in blue) or *pars lateralis* (in red) reverted to LD response in aphids reared under SD conditions. A, antenna; CE, compound eye; TO, triommatidium; OE, oesophagus; NCC, nervi corpori cardiaci; CC, corpus cardiacum; CA, corpus allatum. Bar = 100 µm. Modified from Steel & Lees (1977).

1.3.3 Transcriptomics

The complex (and largely unknown) molecular process behind the photoperiodic induction of sexual reproduction in aphids (and photoperiodic response in general) presumably entails activation and repression of a great number of genes and metabolic pathways. Gene to gene analysis provides useful information, but lacks a broad view of the multiple reactions taking place, involving dozens or hundreds of genes. The emergence of "omic" techniques (proteomics, transcriptomics, genomics, etc.) has provided us with several tools capable of facilitating a broader view of complex processes. The availability of the *A. pisum* genome thanks to an international collaborative effort (The International Aphid Genomics Consortium, 2010) and advances in sequencing technologies, have generated new discoveries in chronobiology using these techniques (Kojima & Green, 2014). Next Generation Sequencing (NGS) studies in mammals have, for example, found thousands of previously unknown circadian controlled genes and even non-protein coding tanscripts which were rhythmically expressed (Kojima & Green, 2014; Yoshitane et al., 2014). Through transcriptional studies it is possible to asses which genes or metabolic pathways are being activated or repressed at a certain moment or under a specific stimulus. Previous studies have performed transcriptional analysis of *A. pisum* to study photoperi-

odic response (Cortés et al., 2008; Le Trionnaire et al., 2009; 2012; Ji et al., 2016). These studies made use of subtractive hybridization, microarrays and RNA-seq experiments respectively. Although microarrays provide invaluable information about gene expression in aphids (Le Trionnaire et al., 2009; 2012), it relies heavily on the current knowledge of the genome, thus excluding potential unknown genes from the study, as arrays are synthesized with known sequences only. Moreover, differences in gene expression are often underrepresented in microarray studies (Kojima & Green, 2014). On the other hand, RNA-seq studies performed by Ji et al. (2016) focused on inter-species divergence, but not in the effect of different photoperiodic regimes. In general terms, the aforementioned transcriptomic studies found similar genes and pathways involved in photoperiodic response. Among the processess and genes found to be affected by photoperiod, some belonged to general categories, such as "metabolism" while a few were more specific. Terms related to housekeeping functions were found, including cytoskeleton and ribosomal protein-related genes (Cortés et al., 2008; Le Trionnaire et al., 2009; 2012). Interestingly, cuticular proteins and pathways related to cuticle synthesis were found to be differentially expressed in all four studies. Also, the neuroendocrine system, hormone production pathways and dopamine and juvenile hormone-related genes were affected by photoperiod. Le Trionnaire an colleagues

(2009) detected differences in the visual system and photoreception in *A.pisum* heads and Ji et al. (2016) found differences in the expression of genes related to food digestion and absorption.

All of these studies, however, covered only one or two photoperiodic conditions. As seen in section 1.1 and Figure 1.3, the aphid photoperiodic response differs depending on the length of the day, producing different proportions of males and sexual females. Male and sexual female induction processes make use of photoperiodic information and initialize pathways that eventually lead to sexual morphs formation. These pathways must have several common elements, as they share the same cue (i.e. photoperiod). Analysis of similarities and differences could be very valuable not only for photoperiodic studies but also to understand polyphenism in aphids.

1.4 Summary of the current knowledge on aphid photoperiodic clocks

After analysing the current knowledge on the mechanisms underlying circadian rhythms and the photoperiodic response in diverse organisms, we will summarize some facts and doubts related specifically with aphids (Figure 1.21).

It is probable that aphids, like other organisms, have a photoperiodic system that can be divided in three basic components (see Figure 1.12):

1. Input
2. Core (Clock and counter)
3. Output

Photoperiodic clock input

In aphids, the light is perceived by photoperiodic photoreceptors in the mother's brain. It is not detected by the mother's eyes nor it is detected directly by the progeny through their mother's abdomen (Lees, 1964; Steel & Lees, 1977) (see section 1.3.1). Steel & Lees (1977) somewhat intuitively proposed that the *pars lateralis* would probably be involved in photoreception (see Figures 1.21 and 1.22).

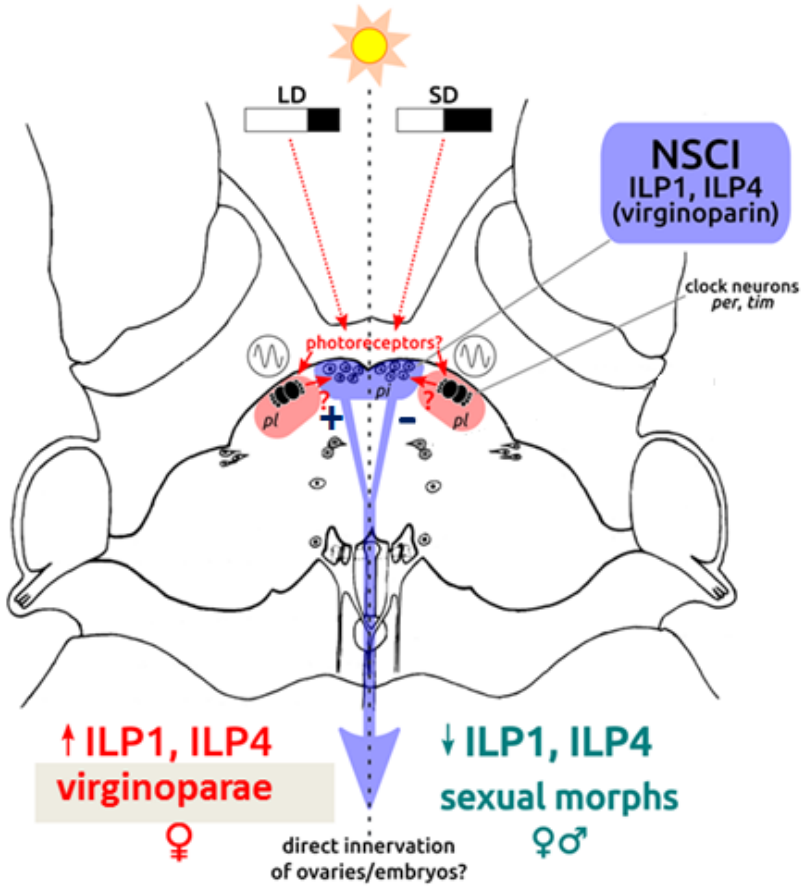


Figure 1.21: Summary of the brain elements involved in aphid photoperiodism. Long days (LD, left half of the graph) or short days (SD, right half) would be perceived through unknown photoreceptors in the *pars lateralis* (*pl*) region. The *pl* contains clock neurons (i.e. the oscillator) expressing relevant clock core genes, such as *period* (*per*) and *timeless* (*tim*). The oscillator is represented by an encircled wavy line. During LDs, insulin-like peptides 1 and 4 (ILP1 and 4) are overexpressed in the group I of neurosecretory cells (NSCI). ILPs have been proposed to be the so called "virginoparin", a substance proposed to be released to the ovaries to induce embryos to develop as virginoparae (Steel, 1976). The double crossed female symbol represents virginoparae.

The main candidate genes to participate in aphid photoreception belong to the opsin and cryptochrome families. These photosensitive proteins could detect light to begin the cascade of reactions that culminates in the photoperiodic response. Opsins have been detected in the neuropile of the anteroventral part of the protocerebrum (Gao et al., 1999). The site of expression of cryptochromes in the aphid brain is still unknown.

Photoperiodic clock core

The aphid core mechanism, which should theoretically include a clock and a counter would also be located in the mother's brain (Steel & Lees, 1977). It is highly likely for the circadian clock to be involved in the integration of daytime information (Nunes & Hardie, 1993). In support of this view, clock genes have been shown to be expressed in the aphid brain and affected by photoperiod (Barberà, 2017). The clock genes *period* and *timeless* are expressed in several groups of neurons (see Figure 1.22), thus defining the aphid's clock neurons (see Figure 1.21). Some of these neurons are named after their position in the protocerebrum: *period* is found in the lamina of the optic lobe (LaN) and in the dorsal and ventral Lateral Neurons (dLN and vLN). *timeless* is expressed in dLN and vLN, but also in a group of large Dorsal Neurons (l-DN). Most relevant, this group

of l-DN is found in the vicinity of the small Dorsal Neurons (s-DN), where both *per* and *tim* are expressed. Both s-DN and l-DN are included in the *pars lateralis*, which is essential to trigger the aphid photoperiodic response, as shown by the microlesion experiments performed by Steel (1978) mentioned above.

Several relevant genes of the circadian clock core were found to be differentially expressed when comparing aphids reared under LD and SD photoperiods: *period*, *timeless*, *clock* and *cycle*, (Barberà et al., 2017).

Photoperiodic clock output

The output pathways consist of neuroendocrine effectors and neurotransmitters from the mother's brain that trigger a cascade of events affecting the developmental pathway of the embryos (Steel & Lees, 1977). These molecules transduce the signal so as to trigger or inhibit the production of sexual morphs depending on information received from the core mechanism. The NSC group I (see Figure 1.20) is necessary in the output pathway (see Figure 1.21).

Some of the candidate molecules to be part of the output pathway in aphids include:

1. Melatonin, which was detected in neurosecretory cells (NSC) in the subesophageal ganglion and thoracic ganglionic mass of

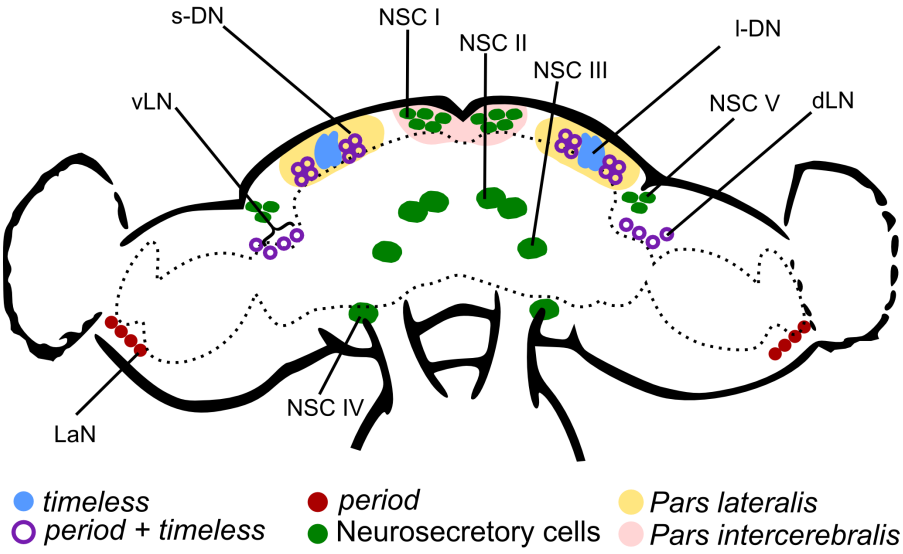


Figure 1.22: Schematic representation of an aphid brain summarising the types of neurons expressing clock and photoperiod related genes. LaN: Lamina Neurons; vLN: ventral Lateral Neurons; s-DN: small Dorsal Neurons; NSC I-IV: Neurosecretory cells I to IV; l-DN: large Dorsal Neuron; dLN: dorsal Lateral Neuron. s-DN and l-DN are located in the *Pars lateralis* (in yellow) and the NSC I are located in the *Pars intercerebralis* (in pink). Adapted from Barberà et al. (2017).

the aphid central nervous system (as well as AANATs, the enzymes that catalyze its production). Melatonin levels increase under SD and is known to be relevant for photoperiodism in vertebrates.

2. PTTH: Although there is no evidence of its involvement in aphid photoperiodic response, it has been demonstrated that PTTH participates in the synchronisation of the brain

clock together with insulin-like peptides (ILPs) in another hemipteran: *Rhodnius prolixus* (Vafopoulou et al., 2014). Barberà & Martínez-Torres (2017) demonstrated that PTTH is produced by the NSC group II in *A. pisum*.

3. Insulin-like peptides: localised in the aphid brain, in the NSC group I. Their expression increases under LDs in holocyclic strains, but not in the anholocyclic strains analysed. Insulin-like peptides and the insulin pathways are involved in the photoperiodic response of other invertebrates (see Figure 1.19). It has been proposed that ILPs are the substance previously termed "virginoparin", which would be released to the ovaries or embryos to stimulate the production of virginoparae (see Figure 1.21).

4. Juvenile Hormone (JH): JH levels are high during LD and mimic long day effects in aphids, as aphids reared under SD revert to parthenogenesis when exposed to JH.

Objectives

With the general objective of contributing to the characterization of the molecular bases governing the photoperiodic response in aphids, our objectives were as follows:

1. The analysis and comparison of transcriptomic profiles through RNA-seq of aphids reared under three different photoperiodic conditions, with special focus on genes related to the input pathway in the photoperiodic response in the pea aphid, *Acyrthosiphon pisum*.

2. The detailed study of two gene families of photosensitive proteins in the pea aphid (*A. pisum*): the opsin and cryptochrome repertoires, covering phylogeny, expression along the day, under different photoperiods and transcript localization in the brain.

These objectives will be set forth in the following chapters. The transcriptomic approach will be detailed in Chapter 2, while the characterization of opsins followed by that of cryptochromes will be addressed in Chapter 3.

Chapter 2

Transcriptomic analysis of photoperiodic induction in *Acyrthosiphon pisum*

2.1 Introduction

Aphids are well known for being remarkable examples of polyphenism (see section 1.1). Despite being clonal organisms, some species produce a surprising number of qualitatively different morphs from the same genetic information. Many of those morphs are induced by environmental conditions such as changes in the photoperiod. As mentioned in the General Introduction (Chapter 1), the pea aphid, (*Acyrthosiphon pisum*) is capable of producing morphologically dis-

tinct sexual morphs (males and sexual females) from genetically identical asexual parthenogenetic mothers (see Figure 1.1). These sexual morphs are born when their mothers sense longer nights, which precede harsh winter conditions. This shortening of the photoperiod is likely detected and processed by the aphid and triggers numerous signaling reactions which will lead, eventually, to a switch in the ontogenic process of the offspring.

In the aphid strain YR2 (see Materials and Methods), a short day photoperiod of 14hs of light and 10hs of darkness (14L:10D) (referred to as SD_{14}) produces males (see Figure 1.3), while further reducing the day length to a 10L:14D photoperiod (SD_{10}) produces sexual females (see Figure 1.3). The long day (LD) condition of 16hs of light doesn't induce the sexual response and results in the production of only parthenogenetic females.

As presentes in Chapter 1, under both response-inducing conditions (SD_{14} and SD_{10}), the shorter photoperiods are detected, and a response is triggered by the coordinator of the response: the aphid central nervous system.

Here we present results of a transcriptomic analysis aimed at comparing levels of expression of genes in the aphid head under three different photoperiods: LD, SD_{14} and SD_{10} which lead to the development of embryos as parthenogenetic females, males and sexual females respectively. Any element shared between the two

SD conditions will be considered to be common to the photoperiodic response, irrespective of the morph produced. Our group focuses in the study of molecular pathways involved in circadian rhythms and the photoperiodic response. In previous works we have analysed genes involved in the core and output elements involved in these pathways (Cortés et al., 2010; Barberà et al., 2017; Barberà et al., 2019).The input pathway, however, was not analysed in depth. Thus, this chapter will not only focus on the general pathways involved in photoperiodic induction in aphids, but will also try to specifically find transcripts that could be candidates to participate in the photosensitive step of the input pathway.

2.2 Objectives

The objective of this chapter is to analyse the transcriptomic profile of pea aphids exposed to 3 different photoperiodic conditions. We aim to identify genes and/or pathways linked to the induction process and to the specific production of males and oviparae (sexual females). The specific objectives are as follows:

1. Obtain the transcriptomes of *A. pisum* exposed to three different light regimes: non-inducing, male-inducing and oviparae-inducing photoperiods.

2. Identify differentially expressed genes and enriched pathways and discuss their putative role in the induction process.

3. Search for transcripts potentially related to photosensitivity in the photoperiodic response.

2.3 Materials and Methods

2.3.1 Aphid rearing

Acyrthosiphon pisum colonies of the strain York Red 2 ("YR2", originally collected in York, UK) were reared at 18^{o}C on *Vicia faba* plants under low density at non-inducing long day (LD) conditions (16 hours of light and 8 of darkness LD, 16L:8D) for two generations to avoid stress signals received from parents (Figure 2.1). Vermiculite was used as substrate for seed germination. Each individual plant containing aphids was covered with a sealed transparent bag. Aphids were synchronized, keeping only those nymphs born on the same day. Forty-five aphids in the L3-L4 nymphal state were transferred in groups of five to the three experimental photoperiodic conditions per triplicate: LD, SD_{14} (14L:10D) and SD_{10} (10L:14D). These aphids were the mothers of our experimental aphids and are referred to as the G0 generation. Thus, individuals born from these mothers (the G1 generation) had been exposed to the experimental photoperiodic conditions before birth. The G1 generation was collected at *zeitgeber* time 9 (ZT9) in groups of 15 individuals and immediately frozen in liquid nitrogen when L3-L4 nymphal state was reached (day 7 in LD and SD_{14}; day 9 in SD_{10}). Some G1 individuals were kept alive to follow up and confirm induction of sexual morph production in the

G2 of SD reared aphids. Each aphid of the follow up G1 population was placed in an individual closed transparent box with two *V. faba* leaflets to facilitate handling of the G2 generation (see Figure 2.1). The reproductive phenotype of aphids born in the G2 generation was registered (i.e. parthenogenetic female, male or sexual female).

2.3.2 RNA extraction and sequencing

Aphid heads were excised in groups of 15 and homogenized with a Pellet Pestle® Motor in TRI reagent®. Direct-zol RNA Miniprep by ZYMO Research was used for total RNA extraction following provider instructions. RNA was quantified using a NanoDrop ND-1000 (Nanodrop Technologies Inc., Wilmington, DE, USA).

RNA sequencing was performed in the Central Service for Experimental Research of the Universitty of Valencia (*Servei Central de Suport a la Investigació Experimental*, SCSIE). RNA quality was assessed with Bioanalyser and samples were enriched in mRNA with the GeneRead Pure mRNA Kit (Qiagen). Nine strand-specific first strand libraries were constructed, which included three replicates for each photoperiodic condition: LD, SD_{14} and SD_{10}. 20M reads 75nt long were obtained per sample in a NextSeq® 550 Illumina sequencer.

The three biological replicates were labelled A1, B1 and C1 for

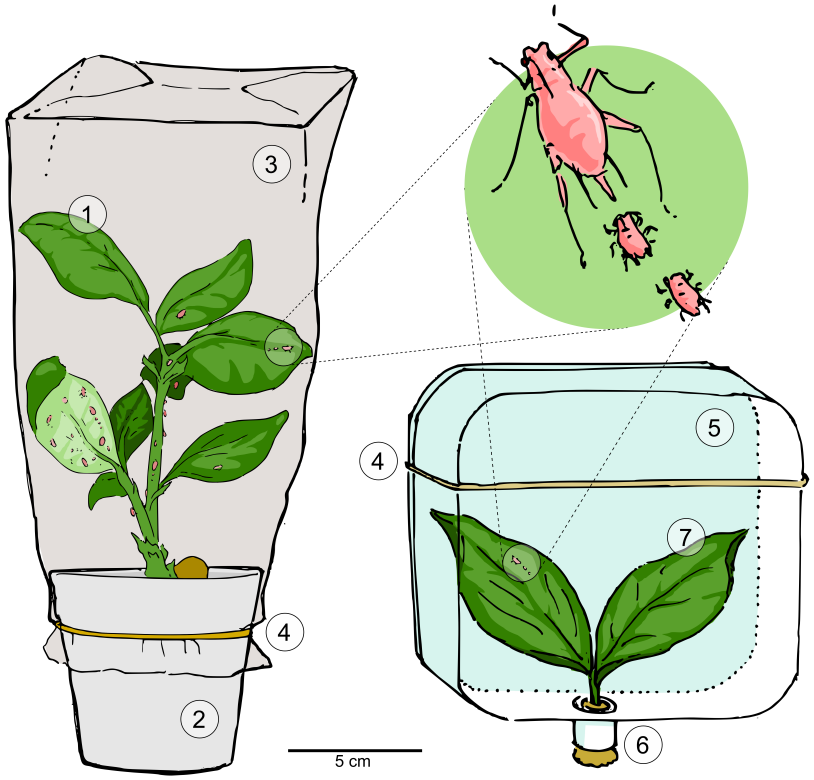


Figure 2.1: Plant and box systems to rear *A. pisum* aphids. *V. faba* plants (1) are grown in plastic pots (2) with vermiculite substrate. The plants are covered with a transparent bag (3) sealed with a rubber band (4). Bags allow air flow through perforations. The plant pot is pierced in the bottom part to allow water absorption from a soaked tray. Plant pots are used for rearing high numbers of aphid. For follow up studies on small numbers of aphids (1-6 individuals) rearing boxes (5) are used. Boxes are made with square petri dishes. A hole is made at one side to introduce a short piece of plastic or rubber tube (6). Two leaflets (7) are introduced in the box, with the peciole in the plastic tube. A piece of sponge (6) holds the peciole, closes the tube and absorbs water from a tray. Rubber bands (4) are used to keep the box shut and to hold it vertically.

LD; A2, B2 and C2 for SD_{10} and A3, B3 and C3 for SD_{14}

2.3.3 Differential expression analysis

The differential expression (DE) analysis was performed after scanning for low quality results and removing sequences below the default scores and sizes allowed in our analysis. The pipeline described below was executed on a Linux command line, using **HISAT2** (Kim et al., 2015), **SAMtools** (Li et al., 2009), **Trim galore!** (Krueger, 2015), **Cuffdiff** (Trapnell et al., 2012) and **CummerBund** (Goff et al., 2019) softwares. When necessary, the R software environment (R Core Team, 2014) was used for statistical computing and graphics. The following steps were followed for indexing the reference genome, transcriptome reads quality trimming, alignment, sorting and to calculate DE. Source code is highlighted with grey background colour. Most commands and options are case sensitive.

1. The fasta file of the pea aphid genome was downloaded from the AphidBase[1] . The genome was indexed using **hisat2** program:

```
hisat2-build -p 15
   assembly2_scaffolds.fasta A_pisum_index
```

[1]http://bipaa.genouest.org/is/aphidbase/iagc/ "Pea aphid genome assembly version 2 scaffolds"

Where:

- hisat2-build = creates an index for a the reference genome "assembly2_scaffolds.fasta".
- -p 15 = launches a specified number (15) of parallel search threads to increase alignment throughput. Optional parameter.[2]
- A_pisum_index = output name of the indexed genome.

2. RNA raw sequences were trimmed and low quality reads were removed using **Trim galore!** software. FastQC reports were generated to asses quality of the high throughput data after trimming:

   For sample UV-A1_S5_.fastq only:

```
trim_galore --length 70 --fastqc --output
   ./trimmed_V UV-A1_S5_.fastq
   1>./trimmed_V/UV-A1_S5_trim.log
   2>./trimmed_V/UV-A1_S5_trim.err
```

   Where:

[2]The "number of threads" parameter is usually optional in many programs and depends on the informatic resources available to the researcher. If not introduced, the program will be executed using one thread only. This parameter only affects time of execution, not the result.

- trim_galore = invokes Trim galore! program to trim and remove low quality sequences
- –length 70 = discard reads that became shorter than 70 bp after trimming
- –fastqc = Run FastQC in the default mode on the FastQ file once trimming is complete
- –output = specifies or creates output location for trimmed sequences: /trimmed_V
- UV-A1_S5_.fastq = file to trim in .fastq format
- 1> = generates a file named "UV-A1_S5_trim.log" in the specified location "./trimmed_V" containing information about the execution fo the program, the stanard output (stdout)
- 2> = if there is an error during the execution, it creates a file in named "UV-A1_S5_trim.err" containing the stadard error (stderr) information in the specified location "./trimmed_V"

For all the samples in the folder:

```
ls -1 *.fastq | (while read filename
do
```

```
trim_galore --length 70 --fastqc --output
   ./trimmed_V $filename
   1>./trimmed_V/$filename.trim.log
   2>./trimmed_V/$filename.trim.err
done)
```

3. RNA "first strand" sequences were aligned to the *A. pisum* genome using **HISAT2**.

   For the "UV-A1_S5__trimmed.fq" sample only:

```
hisat2 -p 15 --dta --rna-strandness R -x
   /index_location/A_pisum_index -U
   UV-A1_S5__trimmed.fq -S
   ./aligned_v/UV-A1_S5.sam
```

   Where:

   - hisat2 = invokes HISAT2 to align sequences to the indexed genome
   - -p 15 = specifies the number of parallel threads (15). Optional.
   - –dta = "downstream transcriptome assembly" report sequence alignment as a SAM (Sequence alignment map) file

- –rna-strandness R = Specifies strand-specific information. "R" stands for the reverse complemented counterpart of a transcript, also known as "first-strand"
- -x = pathway to the index for the reference genome.
- -U = indicates which file will aligned to the genome. In this example "UV-A1_S5__trimmed.fq".
- -S = name of the output SAM file to be created. In this example "UV-A1_S5.sam"

For all the samples in the folder:

```
for i in $( ls *__trimmed.fq ); do

hisat2 -p 15 --dta --rna-strandness R -x
   /index_location/A_pisum_index -U $i -S
   ./aligned_v/$i.sam

done
```

4. Use the **SAMtools** program to convert SAM files to compressed binary versions known as BAM files (.bam). BAM files must be sorted afterwards for the sequences to be organized according to their position in the reference genome, so as to facilitate further analysis.

(a) Convert to bam:

For the " UV-C6_S14__trimmed.fq.sam" sample only:

```
samtools view -bS -@ 15
   UV-C6_S14__trimmed.fq.sam
   >unsorted_bam/UV-C6_unsorted.bam
```

(b) Repeat for all .sam files.

(c) Sort files:

For "UV-A1_unsorted.bam" sample only:

```
samtools sort -@ 5 UV-A1_unsorted.bam
   ./sorted/UV-A1
```

(d) To sort all .bam files:

```
for i in $( ls *_unsorted.bam ); do

samtools sort -@ 15 $i
   ~/your_output_folder/sorted_bam/
   $i.sorted

done
```

Where:

- samtools = invoke SAMtools software

- view -bS = **-b**: produce output in the BAM format.-**S**: input files are in SAM format. The input file is " UV-C6_S14__trimmed.fq.sam"
- -@ 15 = specifies the number of parallel threads (15 or 5). Optional.
- > = indicates the location of the output file generated and its name.
- sort = order unsorted file. Specify input file "UV-A1_unsorted.bam" followed by output folder and file: "./sorted/UV-A1"

5. Use **Cuffdiff** program to calculate DE among samples:

   To compare YR2 LD vs SD_{14} (A1,B1,C1 A3,B3,C3) :

```
cuffdiff -o /sorted_bam/cuffdiff_YR2_
   photoperiod_16_vs_14 -p 7
   /your_GFF_location/AphidBase-OGS-2.0-GFF-
   with-CDS -L YR2_16L,YR2_14L
   UV-A1.bam,UV-B1.bam,UV-C1.bam
   UV-A3.bam,UV-B3.bam,UV-C3.bam
```

   Where:

   - cuffdiff = invoke Cuffdiff software

- -o = specify or create output directory for all generated files. In this example:
  "/sorted_bam/cuffdiff_YR2_photoperiod_16_vs_14"
- -p 7 = specifies the number of parallel threads (7). Optional.
- -L = **L**abel: specify comma-separatd labels for coditions: "YR2_16L,YR2_14L"
- input files: .bam files of replicates must be provided as comma-separated list of one condition followed by a space and the comma-separated list of .bam files of the other conditions. The lists must be in the same order provided in the **-L** option.

To compare YR2 LD vs SD_{10} (A1,B1,C1 A2,B2,C2) using the same parameters as in the comparison above:

```
cuffdiff -o /sorted_bam/cuffdiff_YR2_
    photoperiod_16_vs_10 -p 7
    /your_GFF_location/AphidBase-OGS-2.0-GFF-
    with-CDS -L YR2_16L,YR2_10L
    UV-A1.bam,UV-B1.bam,UV-C1.bam
    UV-A2.bam,UV-B2.bam,UV-C2.bam
```

Cuffdiff produces differential expression files which were sorted

according to $\log_2$ fold changes in expression and p-value. A list of relevant genes was created using the top 10 over and the top 10 under expressed genes (based on $\log_2$ fold change values). The list also includes genes which were found only in one of the conditions, thus being statistically significant but lacking fold change value. Annotations in the differentially expressed gene list were manually searched for specific genes of interest related to photosensitivity, such as opsins and cryptochromes.

ACYPI numbers (*A. pisum* gene ID numbers, see section 1.1) with no function or description assigned were BLAST searched in NCBI to find similar sequences or conserved domains.

Results obtained from cufflinks were explored using the **CummerBund** package (Goff et al., 2019) to obtain a graphical output for visual representation of the data.

Functional annotation analysis with DAVID

Functional annotation analysis was performed online using DAVID 6.8 (Huang et al., 2009) bioinformatics resources (The Database for Annotation, Visualization and Integrated Discovery)[3] to classify groups of similar annotation. ACYPI numbers of differentially expressed transcripts were used as Gene List input in the DAVID database. Parameter values were set at medium stringency levels

[3]https://david.ncifcrf.gov/

in the default option of the Functional Annotation Clustering tool. The *Acyrthosiphon pisum* genome was used as background. Enriched pathways were searched in the KEGG pathways tool included in DAVID. The maximum EASE score[4] value in DAVID, which measures the probability of a gene group of being enriched in the user's sample compared to the reference genome, was set at 0.2, which is slightly less stringent than the default option to allow for more gene groups to be analysed. Enrichment score, or E-score, is calculated with the geometric mean of p-values for all the annotation terms in one group of genes. A minus log transformation is applied, thus the higher the value, more important the role (or enrichment) of the annotation term members. Although values above 1.3 are considered relevant, groups with lower values could be relevant and should be looked into if possible (Huang et al., 2009).

GO term analysis with REVIGO

AgriGO database[5] was used to obtain a list consisting of 8,864 ACYPI numbers and the 27,329 gene ontology terms (GO terms) associated to them. AgriGO provides tools and databases to work with GO terms and is focused in agricultural species (Tian et al., 2017).

[4]A modified (more conservative) Fisher's exact test
[5]http://bioinfo.cau.edu.cn/agriGO/

The differentially expressed ACYPIs obtained using Cuffdiff (see above) was used as query to retrieve their associated GO terms in the list obtained in AgriGO. This was done for the two comparisons performed: LD vs SD_{14} and LD vs SD_{10}.

GO terms from each comparison were used as input in the REVIGO online tool[6] (Supek et al., 2011) to find a representative subset of terms and visualize results. REVIGO clusters similar GO terms according to their semantic similarity[7]. Allowed similarity between terms was set at the default value of 0.7. The whole Uniprot database (the option by default) was used to determine GO terms size (i.e. the percentage of genes annotated with the same term in the database).

[6]http://REVIGO.irb.hr

[7]semantic similarity assesses the degree of relatedness between two entities by the similarity in meaning of their annotations

2.4 Results

2.4.1 Aphid induction follow up

After induction under SD_{14} and SD_{10} conditions for the transcriptomic analysis, the progeny of some G1 individuals was analysed to verify that induction of sexual morphs in the G2 generation was as expected. Table 2.1 registers the number of males, parthenogenetic females and oviparae (sexual females) in the G2 generation given birth during 4-6 days by six individual G1 aphids placed in one of six boxes (labelled A to F). SD_{14} aphids produced males and parthenogenetic females. SD_{10} aphids produced only oviparae. As LD aphids were not exposed to inducing conditions, they continued producing parthenogenetic females exclusively. As that is the default condition of the stock aphids, it is not represented in a table.

2.4.2 Transcriptional data parameters

FastQC

FastQC offers a graphical output to assess sequence quality in the transcriptome. After quality trimming (section 2.3.3), the quality of all the samples was enough to proceed with the analysis. FastQC output was similar to all samples, thus only the results of one sample (YR2, LD sample A1) will be shown here in Figures 2.2 to 2.4. As

Table 2.1: Follow up of G2 generation of aphids from the YR2 strain to confirm induction of male and oviparae production by exposure to SD_{14} (left) and SD_{10} photoperiod (right), both at 18ºC. One individual was placed in each box (A to F) to give birth for 4-6 days and the phenotype of its progeny was registered.

SD_{14}			
Box	Ov.[1]	Male	Part.[2]
A	0	6	18
B	0	6	16
C	0	6	12
D	0	1	9
E	0	2	10
F	0	8	10
Total	0	29	75

SD_{10}			
Box	Ov.[1]	Male	Part.[2]
A	32	0	0
B	15	0	0
C	27	0	0
D	6	0	0
E	0	0	0
F	4	0	0
Total	84	0	0

[1]Oviparae
[2]Parthenogenetic

seen in the FastQC output figures, our RNA-seq samples show high quality per base calling (quality score value > 28) and sequence mean (Phred score > 34). Per base sequence content (Figure 2.3) always presents a similar pattern in RNA-seq experiments, with the first dozen of bases being non random due to hexamer priming during library preparation (priming is never as random as expected) (Hansen et al., 2010). Per sequence GC content corresponds to that of the aphid genome, which lies around 33% (Sabater-Muñoz et al., 2006) (Figure 2.3). As expected after sequence trimming (see section 2.3.3), sequence length is 70-75 nt long (Figure 2.4) and duplication level presents values expected for an RNA-seq experiment (Figure

2.4).

CummerBund graphical output

CummerBund is used to explore data and generate a graphical output. The scatterplot (Figure 2.5) displays the $\log_{10}$(FPKM) values (Fragments Per Kilobase per Million reads) of sequences from the three photoperiods compared. Transcriptomes with similar expression patterns present a distribution of dots (genes) close to the diagonal reference line. The LD condition has expression values more similar to SD_{14} than to SD_{10}, as the dots cloud is tightly wrapped around the reference line. However, in relation to the genes that are not equally expressed in LD and SD_{14}, the higher number of dots above the reference line (see Figure 2.5, left) indicates that more genes have higher expression levels in LD than in SD_{14} (also visible in Figure 2.14, bottom left and right).

Barplots for the top over and underexpressed genes are shown in Figures 2.6, 2.7, 2.8 and 2.9. These graphs also display significantly DE genes which have only been found in one condition, thus no fold change value is available. The expression levels are referred in $\log_{10}$(FPKM), where FPKM (Fragments Per Kilobase per Million reads) is a normalised estimation of gene expression frequently used in RNA-seq analysis. An arbitrary number (1) is added to FPKM (hence FPKM+1) to avoid 0 values when applying a log

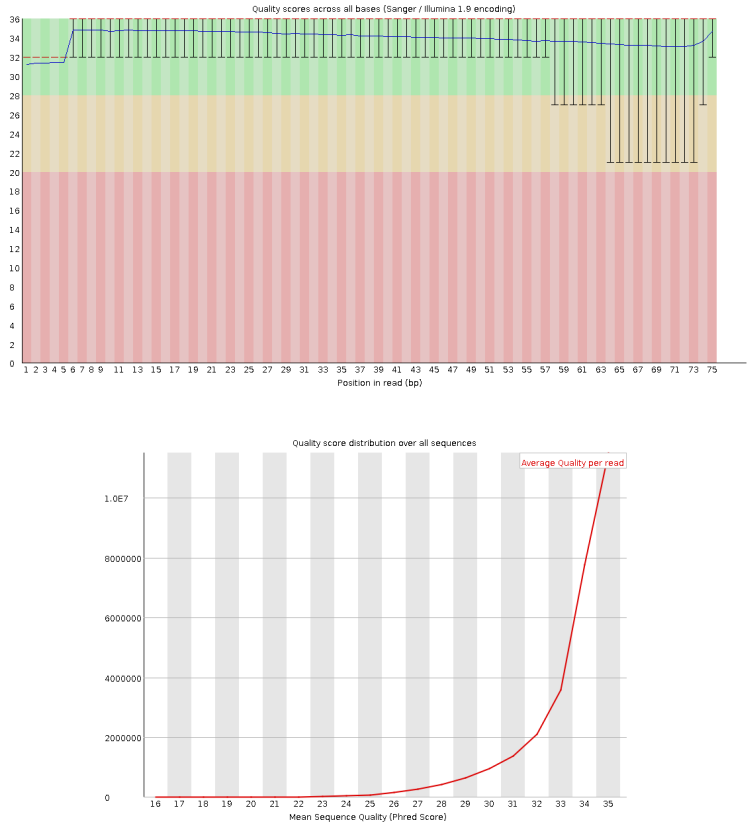


Figure 2.2: FastQC results of YR2, LD sample A1. This result is representative of all the other samples in this experiment **Top**: Quality values across all bases at each position in the FastQ file of the sequences after quality trimming. Good quality base calls are on the green (upper) region of the graph (quality score value > 28). Blue line represents the average quality score for the nucleotide. Drop in quality in the last nucleotides is common in Illumina sequencing. All of our nine samples presented high quality scores in base calling. **Bottom**: Quality score distribution over all the sequences. This graph displays the overall quality of a subset of the sequences present in the transcriptome after quality trimming. A high number of sequences (Y axis) presented good quality values (Phred score $>$ 34).

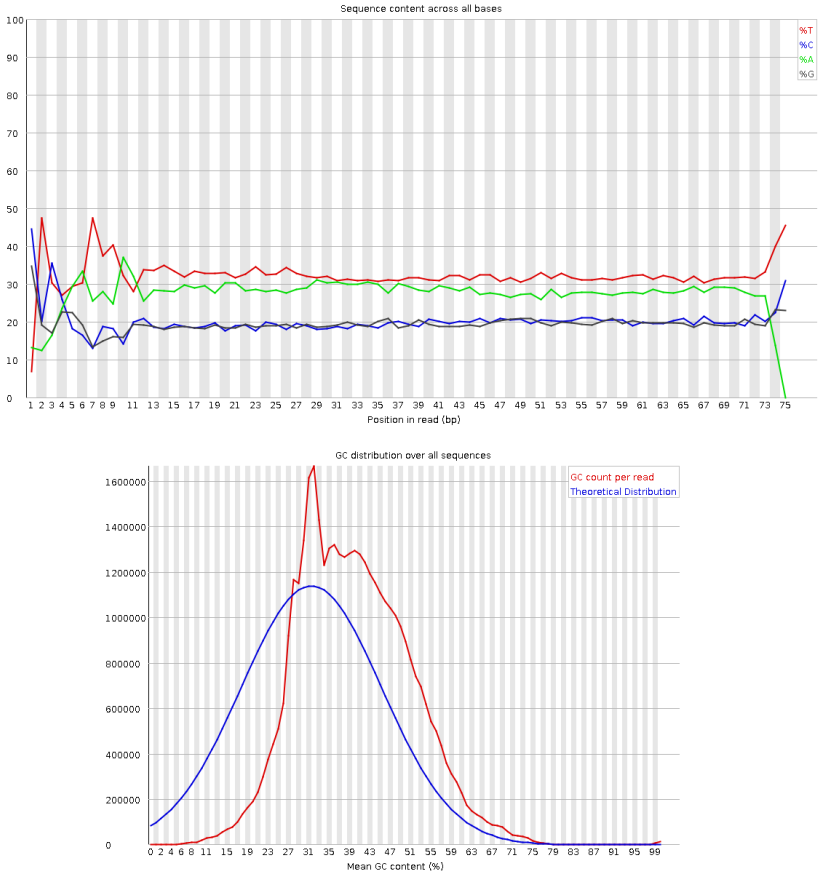


Figure 2.3: FastQC results of YR2, LD sample A1. This result is representative of all the other samples in this experiment **Top**: Per base sequence content across all bases shows a typical distribution of RNA-seq results, with biases in first sequences due to library preparation (Hansen et al., 2010). Y axis displays percentage of each nucleotide in every position. X axis shows the nucleotide position in the sequence. **Bottom**: the GC distribution over all sequences is shown as number of sequences displaying a certain GC content (in %). This result is normal for RNA-seq experiments, as distribution of mean GC content among transcripts might vary, and the central peak is close to the 33% found in the pea aphid genome (Sabater-Muñoz et al., 2006).

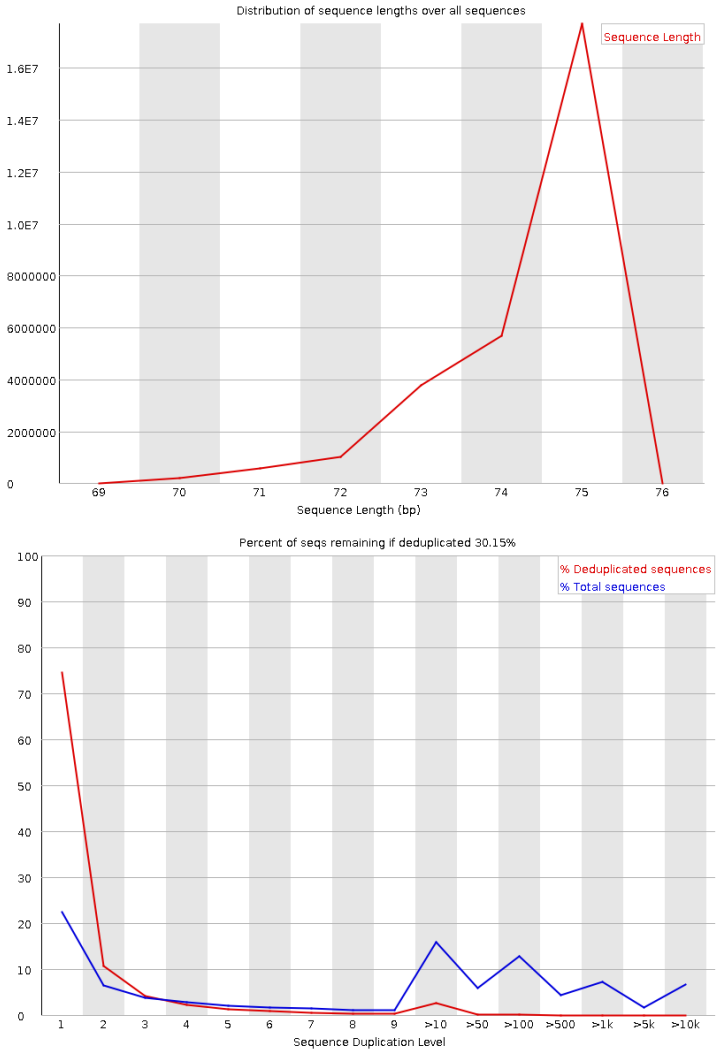


Figure 2.4: FastQC results of YR2, LD sample A1. This result is representative of all the other samples in this experiment. **Top**: The distribution of sequence length shows that most sequences are between 70-75 nt long, which is the length expected after trimming out short sequences. **Bottom**: The sequence duplication level graph shows the percentage of sequncences (Y axis) duplicated an "X" number of times (X axis). All of our samples present a similar result, as RNA-seq experiments are expected to show differential enrichment of many active regions of the genome.

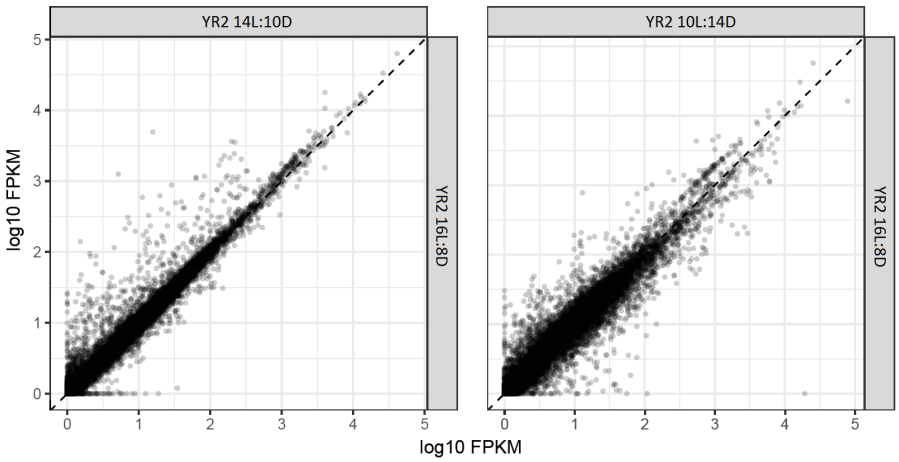


Figure 2.5: Linear scatter plot of FPKM values from RNA-seq analysis of genes expressed under LD, SD_{14} and SD_{10}. FPKM (Fragments Per Kilobase Million) values shown are the means of three replicates for each photoperiod .

transformation for better analysis and representation.

A dendrogram based on Jensen-Shannon distances shows the divergence (where values closer to 0 mean "less diverse" and closer to 1 "more diverse") of the expression profiles of all sample replicates and compares them with samples from a different photoperiod (Figures 2.10 and 2.11). This hierarchical clustering of RNA expression profiles shows that all three samples from the SD_{10} photoperiod form one cluster while the three samples from the LD conditions form another independent cluster (Figure 2.10).

On the other hand, differences between SD_{14} and LD are not as strong, probably because the photoperiodic conditions differ only by two hours. This causes that some SD_{14} expression profiles cluster

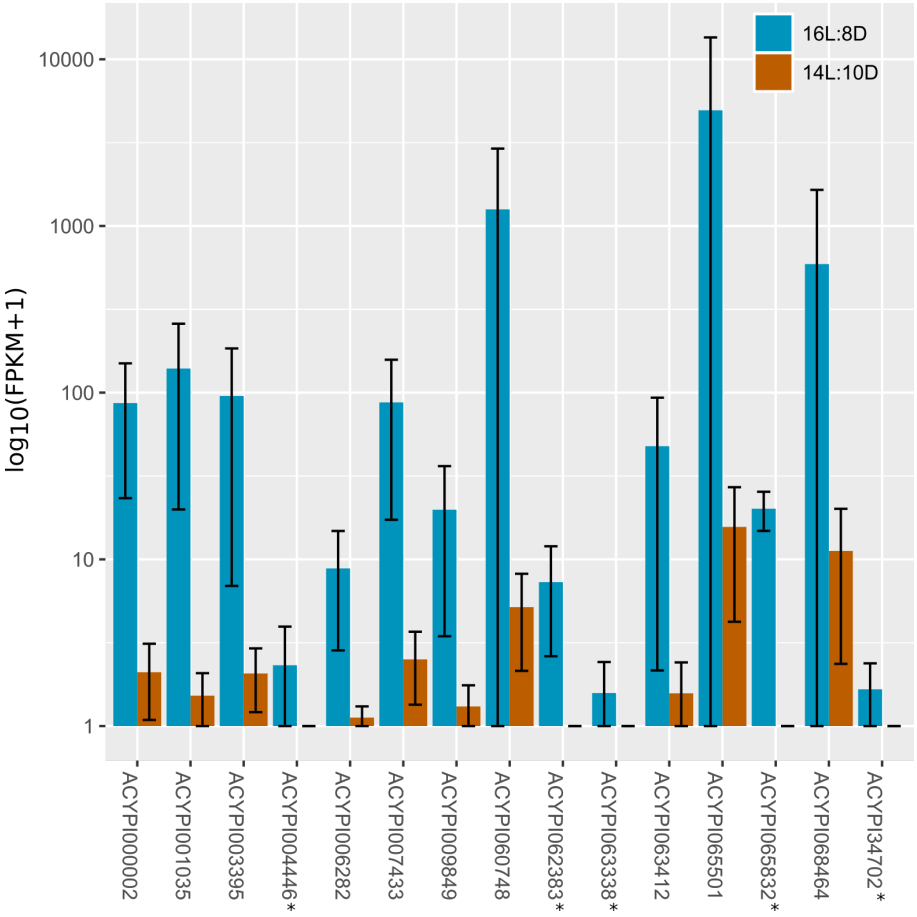


Figure 2.6: Top underexpressed ACYPIs under SD_{14} photoperiod. $\log_{10}$ (FPKM+1) values (Fragments Per Kilobase Million) for the top genes with higher differences in expression are shown. Expression values for genes detected in only one condition (LD or SD_{14}) are also shown (ACYPI with asterisk). Error bars show 95% confidence interval of the mean.

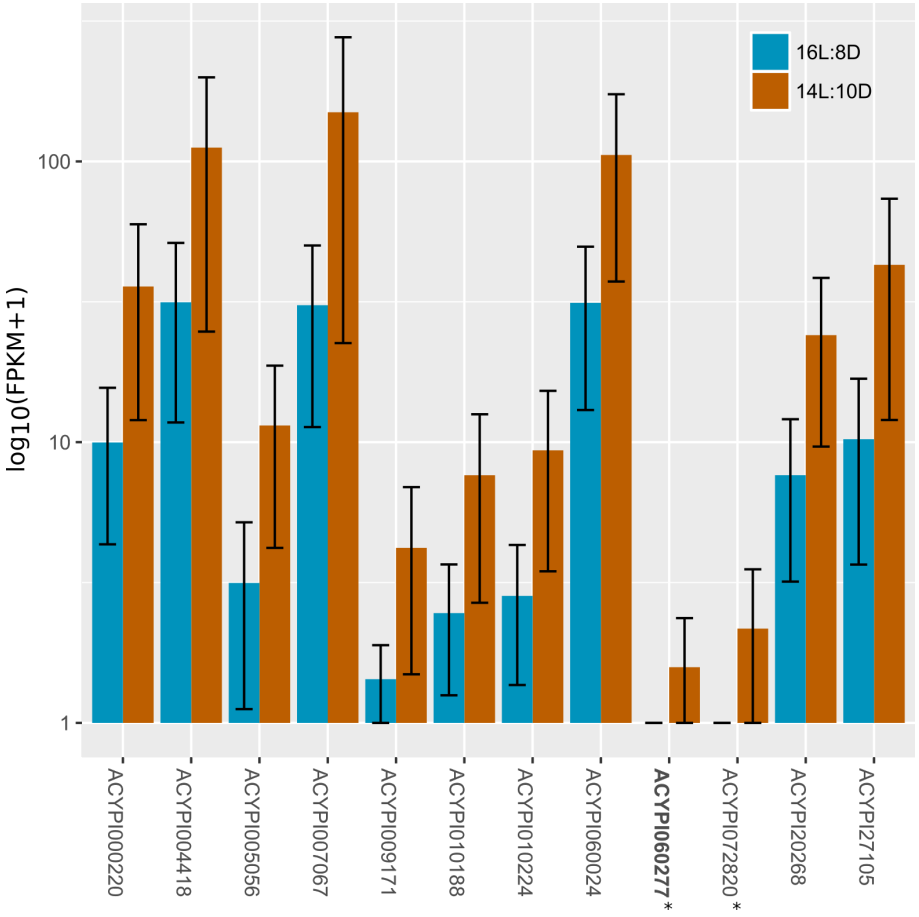


Figure 2.7: Top overexpressed ACYPIs under SD_{14} photoperiod. $\log_{10}$ (FPKM+1) values (Fragments Per Kilobase Million) for the top genes with higher differences in expression are shown. Expression values for genes detected in only one condition (LD or SD_{14}) (ACYPI with asterisk). Error bars show 95% confidence interval of the mean.

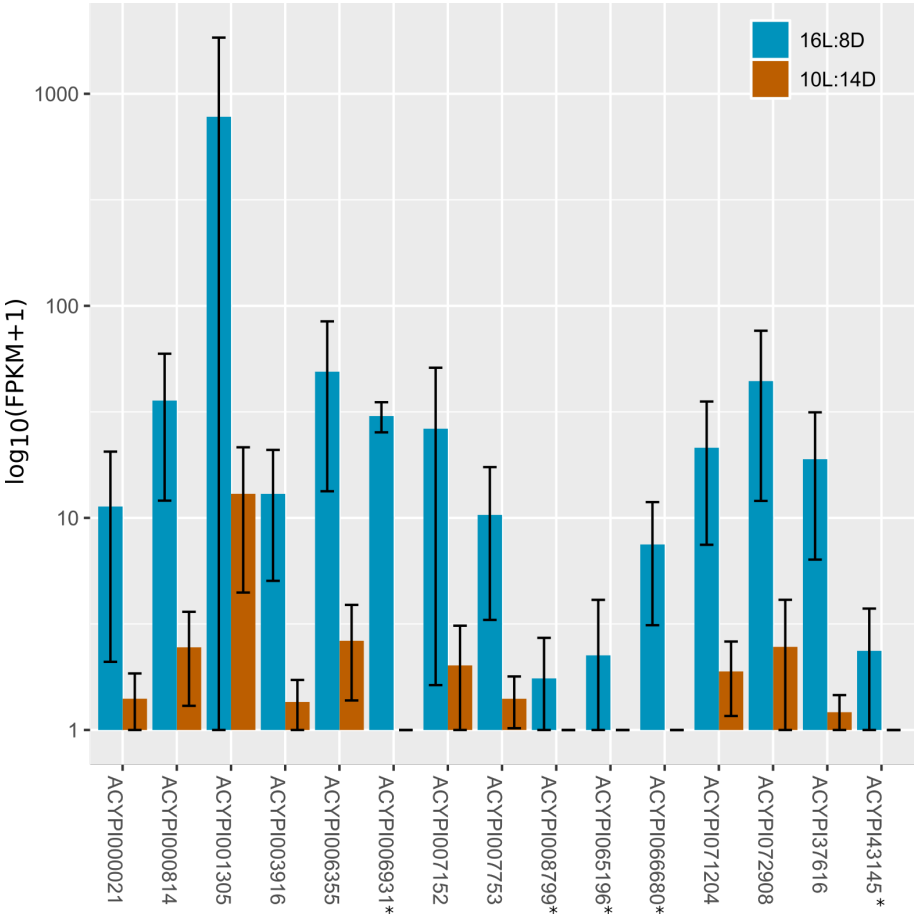


Figure 2.8: Top underexpressed ACYPIs under SD_{10} photoperiod. $\log_{10}$ (FPKM+1) values (Fragments Per Kilobase Million) for the top genes with higher differences in expression are shown. Expression values for genes detected in only one condition (LD or SD_{10}) (ACYPI with asterisk). Error bars show 95% confidence interval of the mean.

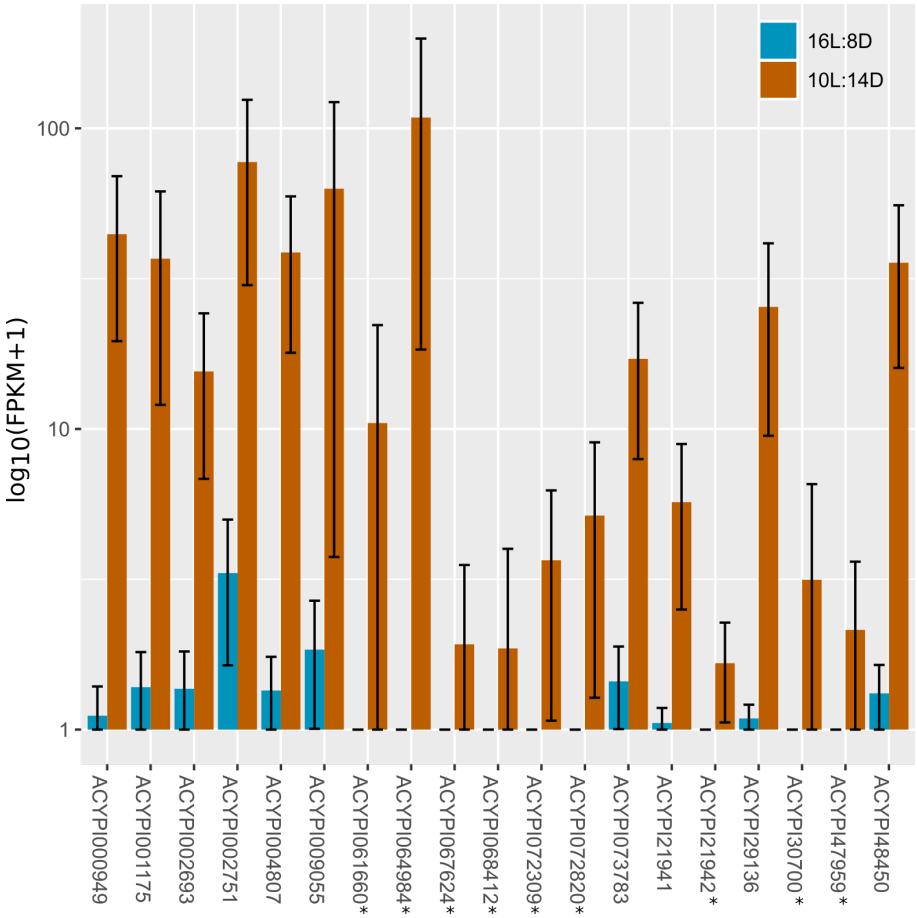


Figure 2.9: Top overexpressed ACYPIs under SD_{10} photoperiod. $\log_{10}$ (FPKM+1) values (Fragments Per Kilobase Million) for the top genes with higher differences in expression are shown. FPKM values for genes present in only one sample are also shown (ACYPI with asterisk). Error bars show 95% confidence interval of the mean.

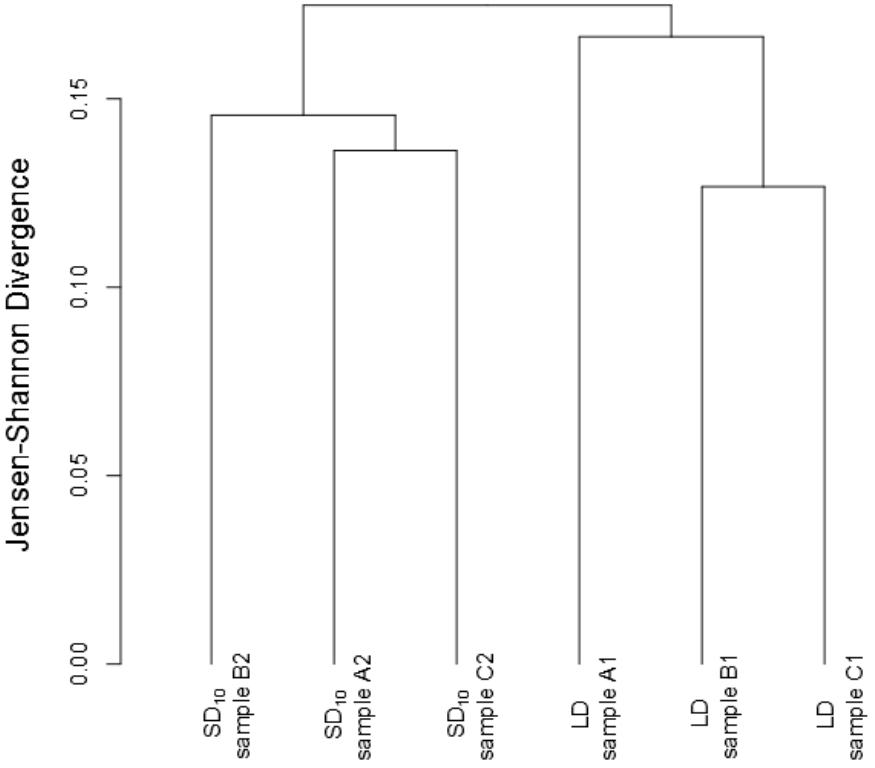


Figure 2.10: Dendrogram displaying Jensen-Shannon distances to show relatedness between expression profiles of all the replicates from the different photoperiodic conditions. This figure shows LD versus SD_{10} photoperiod. Replicates are labelled from A1 to C1 for LD and A2 to C2 for SD_{10}. Samples from the same photoperiod cluster together, resulting in one branch for LD and another one for SD_{10}.

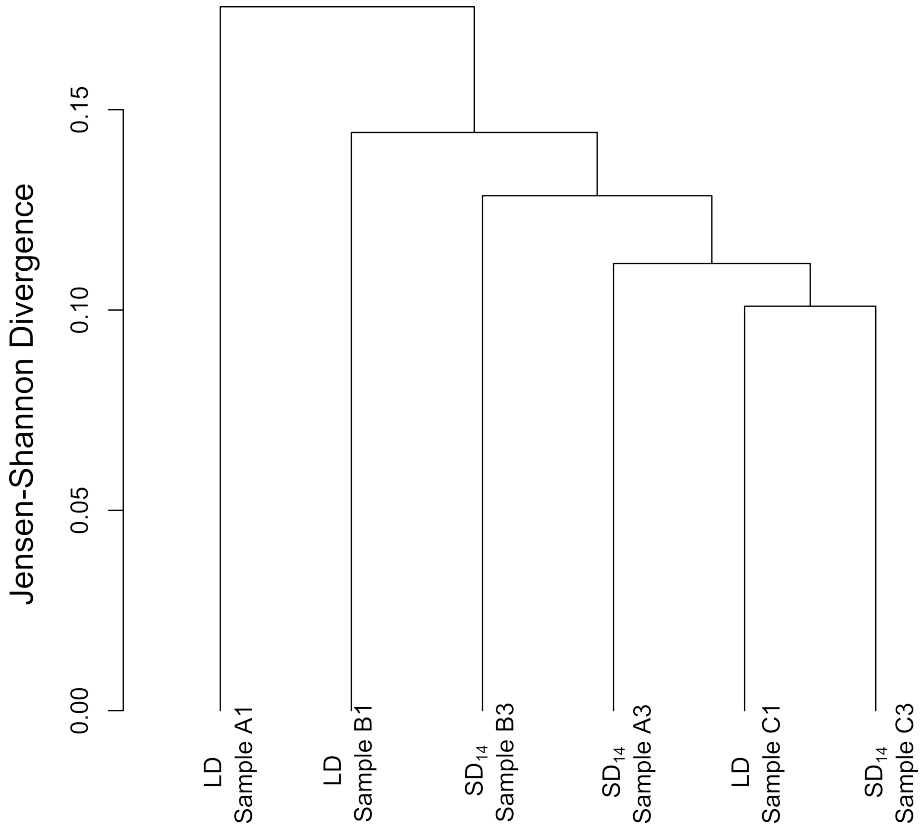


Figure 2.11: Dendrogram displaying Jensen-Shannon distances to show relatedness between expression profiles of all the replicates from the different photoperiodic conditions. This figure shows LD versus SD_{14} photoperiod. Replicates are labelled from A1 to C1 for LD and A3 to C3 for SD_{14}. LD and SD_{14} conditions don't present enough differences in the expression patterns to allow for a clear separation between photoperiods.

together with LD (Figure 2.11).

Volcano plots (Figures 2.12 and 2.13) represent differential expression between genes from two photoperiods (as fold changes), placing magnitude of change as Log_2(fold change) on the X axis and statistical significance (P-value) in the Y axis. These plots highlight the number of differentially expressed genes between conditions. Genes with high Log_2(fold change) appear more distant to the central axis of the graph. Genes with significant p-values appear higher in the graph. Genes with significant p-values are shown as red dots. LD shows less differentially expressed genes when compared to SD_{14} than to SD_{10}, as revealed by the higher density of red dots in Figure 2.13 compared to Figure 2.12.

2.4.3 Differentially expressed genes

Cuffdiff generated two lists of differentially expressed (DE) genes when comparing LD with SD_{14} and SD_{10} (Supplementary Tables S1 and S2). 329 ACYPI entries were found to be DE in SD_{14} and 1072 in SD_{10} when compared with LD conditions.

Under SD_{14}, 87.5% of the DE genes were downregulated in SD (or upregulated in LD), while in SD_{10}, 57%. Without making distinction between up or downregulated genes, 144 DE ACYPIs were shared by both SD photoperiods (Figure 2.14, left). Of all the genes

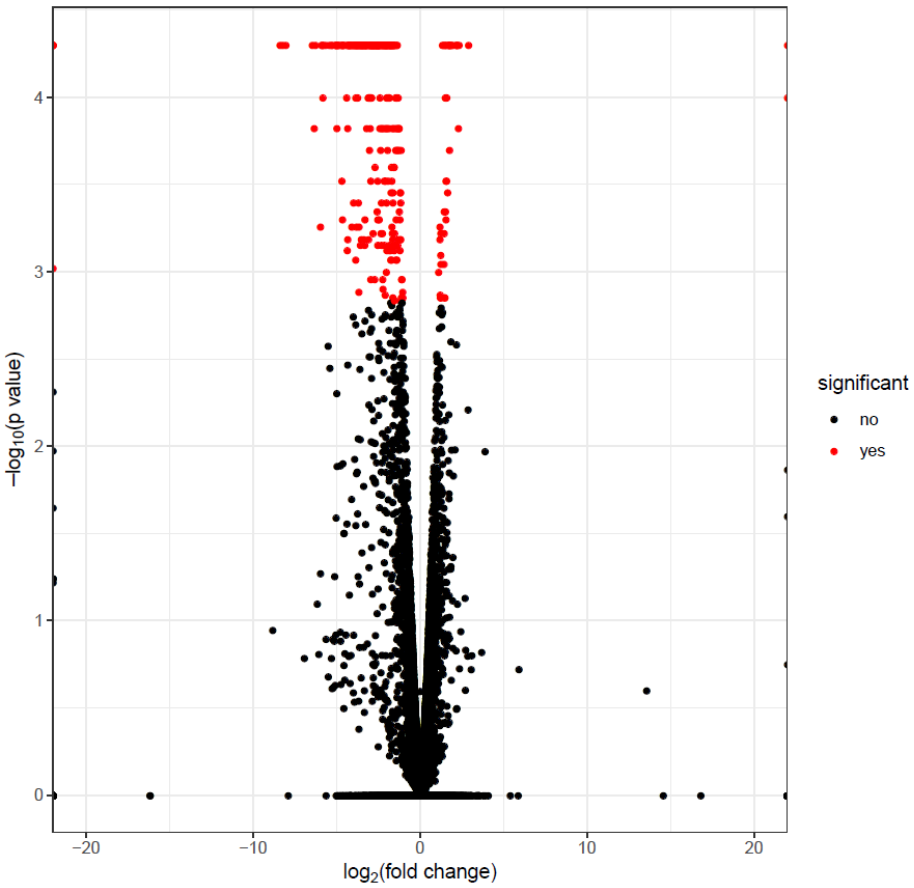


Figure 2.12: Volcano plot: statistical significance (-Log_{10}P-value) versus magnitude of change Log_2(fold change) in differential expression between LD versus SD_{14}. Overexpressed genes under SD_{14} conditions (Log_2(fold change) > 0) fall in the right half of the plot and underexpressed in the left (Log_2(fold change) < 0). Genes with significant differences in expression between conditions are shown as red dots.

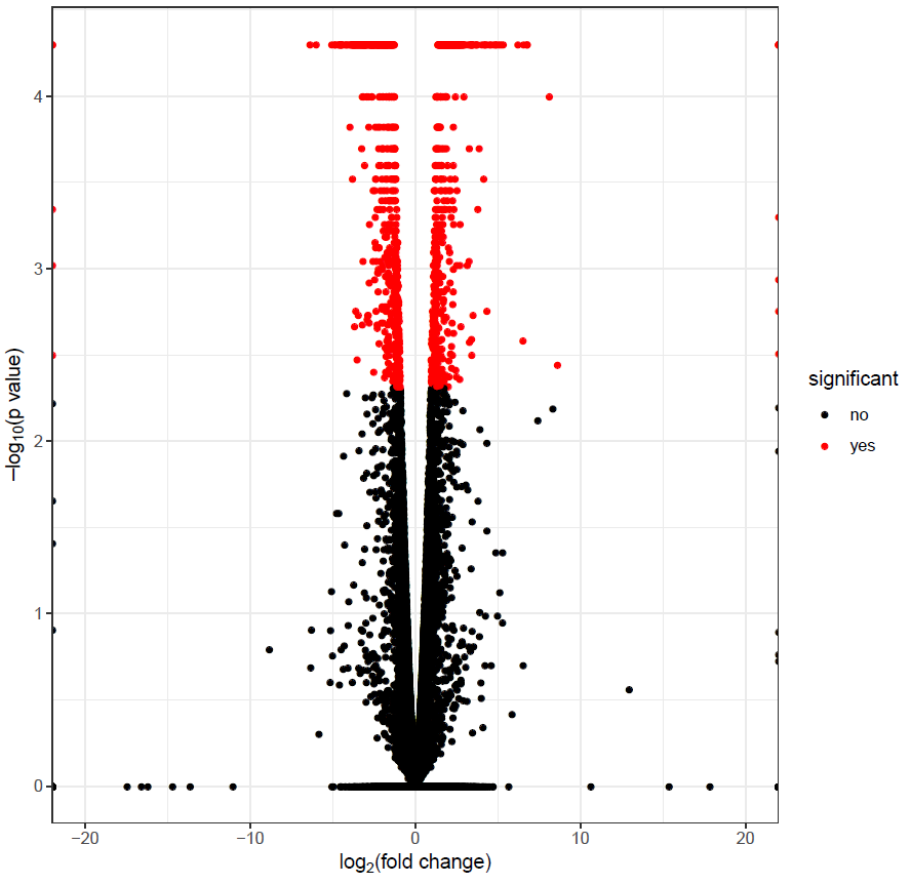


Figure 2.13: Volcano plot: statistical significance (-Log_{10}P-value) versus magnitude of change Log_2(fold change) in differential expression between LD and SD_{10} photoperiodic conditions. Overexpressed genes SD_{10} conditions (Log_2(fold change) > 0) fall in the right half of the plot and underexpressed in the left (Log_2(fold change) < 0). Genes with significant differences in expression between conditions are shown as red dots.

downregulated under SD photoperiods (see Figure 2.14, center), 67 were shared between SD_{14} and SD_{10}. When considering all the genes upregulated under SD, only two ACYPIs where shared between SD_{14} and SD_{10} (Figure 2.14, right). We also calculated which genes were upregulated in one of the SD photoperiods, but downregulated in the other: 62 were upregulated in SD_{10} but downregulated in SD_{14} and in the opposite comparison only 13 genes were upregulated in SD_{14} while downregulated in SD^{10}.

Diferentially expressed genes were ranked according to their $\log_2$(fold change) value. From each comparison, the top underexpressed and top overexpressed genes were extracted for more detailed analysis (Tables 2.2 and 2.3). ACYPIs with no description available were BLAST searched and contrasted with available databases. The Conserved Domain search tool (CD-search) was also used to obtain further information from non- described ACYPIs. In this analysis we included 7 and 14 extra transcripts for LD vs SD_{14} and LD vs SD_{10} respectively found only in one of the conditions, meaning there's no $\log_2$ fold change value available, but presenting statistically significant differential expression. The top transcripts found when comparing LD vs SD_{14} are shown in Figures 2.6 to 2.9 and Tables 2.2 and 2.3. Only one gene, ACYPI072820, was found to be DE under both photoperiodic conditions among the top DE expressed transcripts. No transcripts for this gene were detected un-

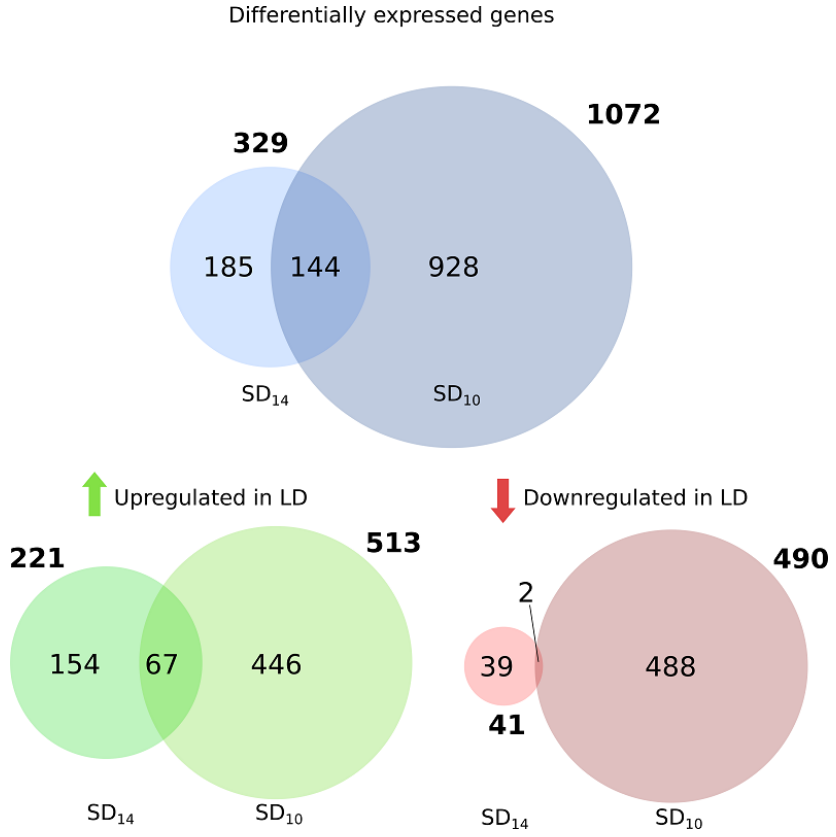


Figure 2.14: Venn diagrams displaying number of shared DE genes (ACYPIs) when comparing long day (LD) photoperiod to two short day photoperiods: SD_{14} (14h light:10h dark) and SD_{10} (10h light:14h dark). **Top**: all shared DE genes, 329 in LD vs SD_{14} and 1072 in LD vs SD_{10}. 185 are exclusive to SD_{14}, 928 are exclusive to SD_{10}. **Bottom left**: ACYPIs upregulated (green arrow) in LD. **Bottom right**: ACYPIs downregulated (red arrow) under LD in both SD photoperiods.

der LD conditions, thus both SD_{14} and SD_{10} show higher expression values. Unfortunately, ACYPI072820 belongs to an uncharacterized protein that yields no BLAST results nor does it contains a known conserved domain

Using CDS tool (Conserved Domain Search tool) we found conserved domains in 4 uncharacterized transcripts: DNA segregation ATPase FtsK/SpoIIIE and related proteins (Cell cycle control, cell division, chromosome partitioning); ZipA superfamily (cell division); C-terminal domain (CTD) superfamily (pre-mRNA processing); and CyaB, thiamine triphosphatase (CYTH)-like Phosphatases[8]. 10 transcripts were uncharacterized and no information could be retrieved from databases or CDS tool.

Table 2.3 shows the top DE genes between LD vs SD_{10}. Using CDS tool with the uncharacterized transcripts we found a cuticular protein of the chitin binding 4 superfamily. Sixteen transcripts were uncharacterized and no information could be retrieved from databases or CDS.

[8](also known as triphosphate tunnel metalloenzyme-like or TTM-like)

Table 2.2: Top differentially expressed genes in LD vs SD_{14}.

	Transcript ID	Gene	log_2 fold change
Downregulated in SD_{14}	ACYPI065501	no hit found – no CD found	-8.4
	ACYPI060748	takeout "to" (circadian controlled by Pdp1)	-8.24
	ACYPI001035	takeout "to" (circadian controlled by Pdp1)	-8.05
	ACYPI003395	takeout-like	-6.47
	ACYPI063412	no hit found – no CD found	-6.35
	ACYPI000002	S1 sucrase	-6.28
	ACYPI006282	venom protease	-5.98
	ACYPI009849	rna-binding protein lin-28	-5.92
	ACYPI068464	no hit found – no CD found	-5.85
	ACYPI007433	lupus la protein sjogren syndrome type b antigen la/ss-b	-5.84
	ACYPI004446	chromobox protein homolog 1	-inf
	ACYPI062383	no hit found – no CD found	-inf
	ACYPI063338	no hit found – no CD found	-inf
	ACYPI065832	no hit found – no CD found	-inf
	ACYPI34702	N-acetyltransferase $ESCO_2$	-inf
Upregulated in SD_{14}	ACYPI060024	uncharacterized protein	1.78
	ACYPI20268	uncharacterized protein	1.8
	ACYPI004418	CDS: DNA segregation ATPase FtsK/SpoIIIE and related proteins (Cell cycle control, cell division, chromosome partitioning)	1.86
	ACYPI000220	Serine proteinase	1.96
	ACYPI27105	CDS: ZipA superfamily (cell division)	2.18
	ACYPI010188	glucose dehydrogenase [acceptor]-like	2.18
	ACYPI010224	double oxidase: two peroxidase domains	2.19
	ACYPI005056	no hit found – no CD found	2.28
	ACYPI007067	CDS: CTD superfamily (pre-mRNA processing)	2.32
	ACYPI009171	Embryonic development factor	2.89
	ACYPI060277	CDS: CYTH-like (also known as triphosphate tunnel metalloenzyme (TTM)-like) Phosphatases	inf
	ACYPI072820	no hit found – no CD found	inf

LD, Long Days; SD_{14}, Short Days with 14L:10D; CDS, result of Conserved Domain Search tool (see text); inf, "infinite", (the transcript was found in only one of the conditions, see text).

Table 2.3: Top differentially expressed genes in LD vs SD_{10}.

	Transcript ID	Gene	log_2 fold change
Downregulated in SD_{10}	ACYPI37616	uncharacterized protein	-6.39
	ACYPI001305	no hit found – no CD found	-6.02
	ACYPI003916	Sp4 Tripsin	-5.07
	ACYPI072908	no hit found – no CD found	-4.88
	ACYPI006355	Cuticiclin-1 like (Component of the cuticles)	-4.87
	ACYPI000021	Cathepsin B (protease)	-4.67
	ACYPI007152	CDS: chiting binding 4 superfamily. Cuticule protein	-4.64
	ACYPI000814	uncharacterized protein	-4.57
	ACYPI007753	maltase H (maltose breakdown)	-4.53
	ACYPI071204	no hit found – unknown CD found	-4.52
	ACYPI006931	Histone H1	-inf
	ACYPI008799	Histone H2A	-inf
	ACYPI065196	no hit found – no CD found	-inf
	ACYPI066680	no hit found – no CD found	-inf
	ACYPI43145	no hit found – unknown CD found	-inf
Upregulated in SD_{10}	ACYPI002751	leucine-rich transmembrane protein	5.04
	ACYPI073783	no hit found – no CD found	5.17
	ACYPI002693	Deoxycytidylate deaminase. Primidine nucleotide metabolic process	5.30
	ACYPI009055	Troponin C	6.19
	ACYPI21941	Peroxidasin (brain development)	6.50
	ACYPI001175	cathepsin B (cysteine-type peptidase activity)	6.54
	ACYPI004807	Serine protease	6.75
	ACYPI48450	Leucine-rich repeats and immunoglobulin-like domains protein 3	6.76
	ACYPI29136	uncharacterized protein- no CD found	8.09
	ACYPI000949	P450 related to CYP4a7	8.59
	ACYPI061660	no hit found – no CD found	inf
	ACYPI064984	no hit found – no CD found	inf
	ACYPI067624	no hit found – no CD found	inf
	ACYPI068412	no hit found – no CD found	inf
	ACYPI072309	no hit found – no CD found	inf
	ACYPI072820	no hit found – no CD found	inf
	ACYPI21942	GATA zinc finger domain-containing protein 4 – like	inf
	ACYPI30700	Craniofacial development protein 2 /	inf
	ACYPI47959	Ubiquitin-like protein	inf

LD, Long Days; SD_{10}: Short Days with 10L:14D; CDS, result of Conserved Domain Search tool (see text); inf, "infinite", (the transcript was found in only one of the conditions, see text).

2.4.4 DAVID functional clustering and KEGG pathways

DAVID functional clustering

DAVID functional clustering tool classified the identified DE ACYPI IDs in several groups according to the similarities in their annotation profiles. In SD_{10}, 690 out of 1072 ACYPIs were included in the analysis (64.34%). In SD_{14}, 204 out of 329 ACYPIs were included in the analysis (62%). See supplementary Tables S3 and S4 for detailed clustering output.

More clusters were found in SD_{10} than in SD_{14} (48 and 15 respectively). 13 of the clusters in SD_{14} were also present in SD_{10}. The acyltransferase activity and the major facilitator superfamily domain were exclusive to SD_{14}, which might indicate a relation with the generation of male progeny.

The 13 clusters shared between both comparisons include terms related with transmembrane proteins, serine protease activity, peptidase C1A, fatty-acyl-CoA reductase, Glucose-methanol -choline oxidoreductase, Pyridoxal phosphate-dependent transferase, CRAL/ TRIO domain[9], Glycoside hydrolase, ABC transporters, oxidoreductase/ Heme binding (CYP P450), Kelch-like protein and Zinc fingers (Table 2.4). Clusters 14 and 15 in SD_{14}, the Zinc Finger

[9]Cellular retinaldehyde (CRAL)-binding protein and Triple functional domain protein (TRIO) guanine exchange factor.

Table 2.4: Gene clusters enriched both under SD_{14} and SD_{10} inducing photoperiods detected with DAVID tools.

Cluster	E-Score SD_{14}	E-Score SD_{10}
1. Transmembrane proteins	3.57841	4.75635
2. Serine protease activity	2.86369	3.32068
3. Peptidase C1A	2.22813	1.13206
4. fatty-acyl-CoA reductase, peroxisome	2.01609	1.23764
5. Glucose-methanol-choline oxidoreductase	1.41751	0.73715
6. Pyridoxal phosphate-dependent transferase	1.04272	1.02603
7. CRAL/TRIO domain	0.93757	1.50055
8. Glycoside hydrolase	0.83747	0.45182
9. ABC transporters	0.70387	2.56685
10. oxidoreductase/Heme binding (CYP 450)	0.55729	1.82813
11. Kelch-like protein	0.37618	0.34355
12. Zinc finger	0.27088	0.39166

E-Score: Enrichment Score. See text for details.

domain clusters, were treated as one in Table 2.4, thus the table displays 12 common clusters only.

KEGG pathways

Transcripts were searched for the presence of enriched pathways with the KEGG tool integrated in DAVID (Tables 2.5, 2.6 and 2.7).

Of the differentially expressed genes under SD_{14}, only 15.2% were included in KEGG analysis. When LD aphids were compared against the male producing photoperiod (SD_{14}), we found upregulation in LD in pathways related to "peroxisomes", "ABC transporters", "lysosome", "tryptophan metabolism", and "synthesis and degradation of ketone bodies" (Table 2.6). When considering less stringent conditions (Ease score conditions of 0.2 in stead of 0.1) we found

Table 2.5: KEGG results in LD vs SD_{14} conditions for pathways upregulated in LD.

Term	EASE[1] threshold	Fold Enrichment	Count[2]
Peroxisome	0.1	8.3640	6
ABC transporters	0.1	16.5358	4
Lysosome	0.1	4.8340	5
Tryptophan metabolism	0.1	12.4018	3
Synthesis and degradation of ketone bodies	0.1	26.6410	2
Tyrosine metabolism	0.2	14.1040	2
Butanoate metabolism	0.2	14.1040	2
Phototransduction - fly	0.2	11.4175	2

[1]EASE threshold: Ease score threshold that includes a term.

[2]Count: number of genes in the list related to a term (>2).

upregulation in LD in the categories "butanoate metabolism", "tyrosine metabolism" and "phototransduction in flies" (see Table 2.6). DAVID didn't find downregulated pathways in LD when compared to SD_{14}.

Regarding genes DE in sexual female producing conditions (SD_{10}), 26.8% of the ACYPI IDs were associated with an enriched KEGG pathway when compared with LD aphids. Upregulation was found in genes related to ribosome biogenesis in eukaryotes, DNA replication and RNA transport. Under EASE score = 0.2, other pathways were detected: FoxO signaling pathway, ABC transporters, Lysosome and Spliceosome. Other genes were found to be downregulated, such as those involved in the categories named

Table 2.6: KEGG pathways results in LD vs SD_{10} for pathways upregulated in LD.

Term	EASE[1] threshold	Fold Enrichment	Count[2]
Ribosome biogenesis in eukaryotes	0.1	4.3612	17
DNA replication	0.1	4.3291	9
RNA transport	0.1	2.1422	18
FoxO signaling pathway	0.2	2.3480	6
ABC transporters	0.2	3.1846	4
Lysosome	0.2	1.6758	9
Spliceosome	0.2	1.6033	10

[1]EASE threshold: Ease score threshold that includes a term.

[2]Count: number of genes in the list related to a term (>2).

Table 2.7: KEGG pathways results in LD vs SD_{10} conditions for pathways downregulated in LD.

Term	EASE threshold	Fold Enrichment	Count
Metabolic pathways	0.1	1.4084	23
Insulin resistance	0.1	4.6178	4
Ascorbate and aldarate metabolism	0.1	3.5121	4
Glycosphingolipid biosynth. (ganglio series)	0.2	4.6754	3
Peroxisome	0.2	2.8995	4

Count: number of genes in the list related to a term (>2).
EASE threshold: Ease score threshold that includes a term.

Metabolic pathways, Insulin resistance and Ascorbate and aldarate metabolism. Glycosphingolipid biosynthesis - ganglio series and peroxisome related genes were detected with 0.2 EASE score values.

2.4.5 REVIGO analysis

As mentioned in section 2.4.3, Cuffdiff generated a list of DE genes, including up and down-regulated genes. 329 ACYPI entries were found to be DE in SD_{14} and 1072 in SD_{10} respectively when compared to LD. We searched the AgriGO database entries for GO terms associated to those ACYPIs. We obtained one GO term list for both inducing photoperiods (SD_{14} and SD_{10}). The two lists were introduced individually in the REVIGO online tool to obtain a list of non-redundant and relevant GO terms (considering their semantic relation with other terms and uniqueness, see below).

In the case of SD_{14} photoperiod, 123 out of the 329 DE ACYPI entries were matched to 398 GO terms from AgriGO. REVIGO reduced the 398 GO terms to 130 (Supplementary Table S5).

Under SD_{10} conditions, 455 out of the 1072 DE ACYPI entries were matched to 2028 GO terms from AgriGO. REVIGO reduced the list to 274 GO terms (Supplementary Table S6).

Results lists were searched for terms directly related to photoperiodism and photosensitivity and were summarized in Tables 2.8 and 2.9. Genes included in the summary are related with vision, photosensitivity, eye development, hormones, circadian rhythms, photoperiod and brain function. Sex-specific genes were also found and included in the list.

REVIGO provides a frequency value, which indicates the occurrence of the GO term in the REVIGO database used as background. Also, it provides a uniqueness value. It measures if a given term is an outlier when compared semantically to the input list (Supek et al., 2011). Semantic similarity measures the degree of relatedness between two entities by the similarity in meaning of their annotations (Pesquita et al., 2009). When the input file is a list of GO terms with no other numerical value (as is our case) REVIGO prioritizes terms with higher 'uniqueness' (Supek et al., 2011).

Table 2.8: REVIGO selected results of GO terms differentially expressed when comparing LD vs SD_{14} conditions. See text for details

Term ID	Description	Freq.[1]	Uniq.[2]
GO:0007632	visual behavior	0,012%	0,616
GO:0048066	developmental pigmentation	0,020%	0,907
GO:0008057	eye pigment granule organization	0,001%	0,824
GO:0009416	response to light stimulus	0,157%	0,860
GO:0045463	R8 cell development	0,00%	0,683
GO:0016318	ommatidial rotation	0,002%	0,674
GO:0045467	R7 cell development	0.001%	0,666
GO:0031290	retinal ganglion cell axon guidance	0.006%	0,545
GO:0007455	eye-antennal disc morphogenesis	0,002%	0,696
GO:0007617	mating behavior	0,012%	0,674
GO:0008049	male courtship behavior	0,004%	0,688
GO:0048511	rhythmic process	0,077%	0,985
GO:0042416	dopamine biosynthetic process	0,003%	0,906
GO:0042053	regulation of dopamine metabolic process	0,003%	0,878
GO:0048009	insulin-like growth factor receptor signaling pathway	0,007%	0,838
GO:0006950	response to stress	4,575%	0,878
GO:0048545	response to steroid hormone	0,118%	0,865

[1]Freq., Frequency. See text for details.

[2]Uniq., Uniqueness. See text for details.

Table 2.9: REVIGO selected results of GO terms differentially expressed when comparing LD vs SD_{10} conditions. See text for details.

Term ID	Description	Freq.[1]	Uniq.[2]
GO:0008057	eye pigment granule organization	0,001%	0,838
GO:0007458	progression of morphogenetic furrow (compound eye morphogenesis)	0,000%	0,581
GO:0016318	ommatidial rotation	0,002%	0,5
GO:0046667	compound eye retinal cell programmed cell death	0,001%	0,526
GO:0007464	R3/R4 cell fate commitment	0,001%	0,514
GO:0045468	regulation of R8 cell spacing in compound eye	0,000%	0,578
GO:0042683	negative regulation of compound eye cone cell fate specification	0,000%	0,535
GO:0007455	eye-antennal disc morphogenesis	0,002%	0,512
GO:0007632	visual behavior	0,012%	0,559
GO:0045463	R8 cell development	0,000%	0,555
GO:0045467	R7 cell development	0,001%	0,534
GO:0045470	R8 cell-mediated photoreceptor organization	0,000%	0,548
GO:0007622	rhythmic behavior	0,012%	0,947
GO:0007617	mating behavior	0,012%	0,569
GO:0042745	circadian sleep/wake cycle	0,005%	0,947
GO:0006584	catecholamine metabolic process	0,001%	0,944
GO:0042416	dopamine biosynthetic process	0,003%	0,941
GO:0042427	serotonin biosynthetic process	0,002%	0,943
GO:0015842	aminergic neurotransmitter loading into synaptic vesicle	0,001%	0,895
GO:0015872	dopamine transport	0,007%	0,896
GO:0030431	sleep	0,012%	0,612
GO:0048009	insulin-like growth factor receptor signaling pathway	0,007%	0,847
GO:0016319	mushroom body development	0,004%	0,568
GO:0046928	regulation of neurotransmitter secretion	0,012%	0,783

[1]Freq., Frequency. See text for details.

[2]Uniq., Uniqueness. See text for details.

2.4.6 Candidate genes involved in photosensitivity: opsins and cryptochromes

Opsins and cryptochromes are light sensitive proteins frequently studied for their relation with circadian rhythms and photoperiodic responses (or behaviors). We have searched specific ACYPI accession numbers associated to opsin and cryptochrome genes (described in detail in Chapter 3). We extracted expression values for those ACYPIs to analyze their expression under SD14 and SD10 when compared to LD (see Figure 2.15 and supplementary Tables S7 and S8). Regarding the cryptochrome gene cluster, ACYPI005768 (*Cry 1* homolog) and ACYPI004197 (*Cry 2* homolog) were found in our transcriptional study (Figure 2.15).

Only one ACYPI corresponding to a sequence encoding a gene member of the opsin family showed significant differential expression in the transcriptomics experiment: Ap-arthropsin is significantly overexpressed under SD_{10} conditions. Regarding the other genes, although not significantly, the vast majority seems to increase expression under both short day conditions (SD_{10} and SD_{14}) (see Figure 2.15). That difference is even higher in the shortest photoperiod (SD_{10}). Only Ap-SWO1 (ACYPI001006) and Cry 1 (ACYPI005768) were underexpressed in SD_{14} conditions. This difference was also not significant.

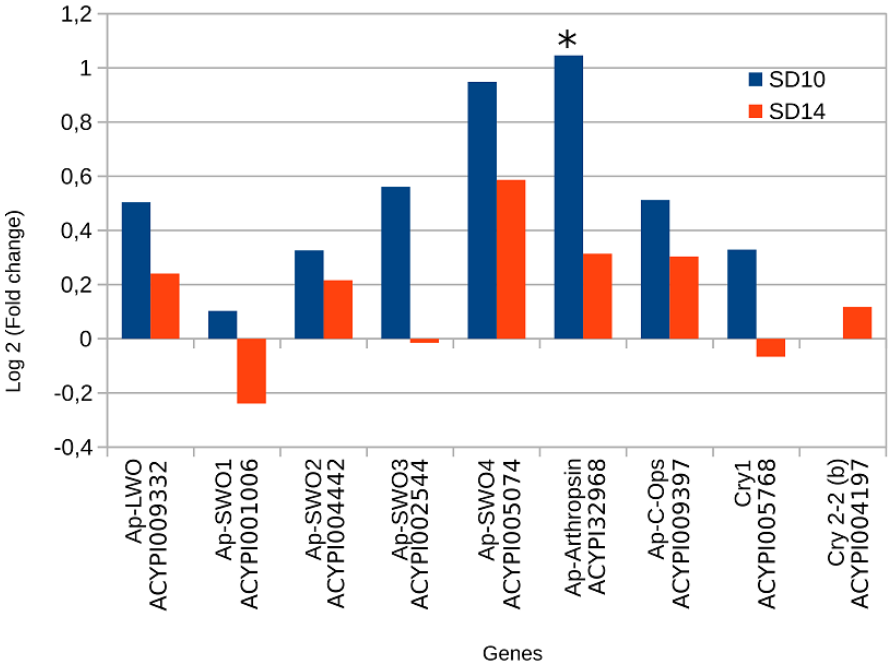


Figure 2.15: Barplot representing Log_2(fold change) of opsin and cryptochrome genes exposed to short day photoperiods (SD_{14} and SD_{10}) versus LD. Asterisk shows significant difference q-value<0.05 (adjusted p-value, taking into account the false discovery rate, FDR). Data from Tables S7 and S8 in the supplementary Material.

2.5 Discussion

Differential gene expression analysis shows that, compared to LD photoperiod, more genes are differentially expressed as the photoperiod shortens. That is, when comparing long day photoperiods with short photoperiods, the shorter the photoperiod, the more differentially expressed genes (see Figure 2.14). This might be due to the more extreme conditions perceived by the putative photoperiodic potoreceptor and response mechanism, as shorter days (or longer nights) may be interpreted as an imminent arrival of winter.

2.5.1 Gene clusters differentially expressed under SD conditions

DAVID clustering analysis provided clusters of DE genes under the three different photoperiods analysed (Table 2.4). Clustering allows for easier interpretation of biological meaning in differential expression by grouping highly related genes. Compared to LD, less genes were DE under male inducing conditions (SD_{14}), thus generating less clusters (15) than under the sexual females inducing condition (SD_{10}, 48 clusters). Male inducing conditions (SD_{14}) are more similar to the LD photoperiod, with a smaller difference in day length (only two hours shorter). In fact, under SD_{14} the parthenogenetic-producing pathway is still active, as this type of

female is also produced along with the males (see Figure 1.3). This subtler change probably has less impact on the organism, thus generating less number of DE genes and clusters.

Moreover, out of the 15 clusters found in SD_{14}, 12 were also found to be DE in SD_{10} conditions. These shared clusters are probably part of hypothetical common pathways involved in the photoperiodic response, regardless of the length of the photoperiod involved. Among the shared clusters, the largest one (cluster 1, see Table 2.4) includes "transmembrane proteins", a rather general term. It is likely that several molecules and cellular products are imported, exported and detected to coordinate the photoperiodic response, a complex phenomenon that probably requires intense intercellular communication and sensitivity (Hegazi et al., 2019). This highly interactive process probably involves multiple neuropeptides and neurotransmitters (Hegazi et al., 2019). If this is the case, the transcription levels of diverse genes related to transmembrane proteins is altered during photoperiodic induction.

Serine proteases (cluster 2, see Table 2.4) and serine protease inhibitors are mainly studied for their role in digestive processes, immune response or development (Kanost & Clem, 2017). Interestingly, they are also localized in the brain of mammals and some species of locust (Harding et al., 1995; Clynen et al., 2005) and might be involved in serine protease-mediated neuropeptide precur-

sor processing (Clynen et al., 2005). Neuropeptides and peptide hormones are thought to play relevant roles in the photoperiodic counter and the photoperiodic endocrine effectors (output of the photoperiodic clock) (Takeda & Skopik, 1997; Danks, 2003). Thus, serine proteases might be involved in processing neuropeptide or peptide hormones released to the haemolymph during the aphid photoperiodic response.

Cystein-type peptidase activity (C1A) (cluster 3, see Table 2.4) was found to be DE in LD vs. SD aphids. It is related to venomous substances in insects, brain remodelling and digestion. Some aphid species, like *Tuberaphis styraci*, have a soldier caste that paralyses and kills attackers by injecting venom through the stylet. The main component of this venom is an enzyme of the cathepsin B family with cystein-type peptidase activity (Kutsukake et al., 2004). *A.pisum*, however, doesn't produce soldiers, thus we discard a use of this enzymatic activity in colony defense. Brain remodelling in the flesh fly *Sarcophaga peregrina* development involves cystein peptidase activity in metamorphosis (Fujii-Taira et al., 2000). Although aphids do not go through metamorphosis, nymphs do undergo brain remodelling as they advance through the different stages of moulting. This brain development might be related with necessary connections to trigger the photoperiodic response, such as that needed to carry the message from the mothers brains to the embryos. Cystein

peptidases make an important part of the aphid digestive enzymes (Cristofoletti et al., 2003) so they might be carried over in the dissection process. If that is the case, we can't discard variability in digestive enzyme production due to food quality variation caused by effect of photoperiod on the *V. faba* plants.

Fatty acyl-CoA reductases (FARs) (cluster 4, see Table 2.4) are tightly related to pheromone and cuticle biosynthesis (Teerawanichpan et al., 2010, Lienard et al., 2010, Finet et al., 2019). Although some insects, like the wasp *Vespula squamosa* do seem to produce pheromones in the head (Landolt et al., 1999), there is no evidence to suggest this could be the case in *A.pisum*. In relation with cuticle biosynthesis and biological rhythms, it is known that in many insect species, cuticle is thickened every day by depositing a pair of chitin layers, each with a different orientation (Neville, 1983). Cuticle deposition rhythm in the bean bug *Riptortus pedestris* and in *Drosophila melanogaster* is regulated by a circadian clock. Also, previous studies in aphids have detected that seasonal photoperiodism affects cuticular proteins (Le Trionnaire et al., 2007; Cortés et al., 2010). In fact, in the study by Le Trionnaire et al. (2007), the majority of the significantly regulated transcripts found were different cuticular proteins, suggesting some cuticle remodelling resulting as part of the photoperiodic response.

It is possible that enzymes related to cuticle modification (such

as FARs) could be affected in our experiment as sampled aphids were close to moulting. Moreover, different photoperiods delay or accelerate moulting time, thus providing a possible explanation to FAR differential gene expression. Le Trionnaire et al. (2007), however, demonstrated that some cuticular proteins are differentially expressed by the exposure to different photoperiods but not by moulting. They proposed that variations in the cuticle due to photoperiodism might cause differences in the perception of the quantity of light.

Pyridoxal phosphate-dependant transferase (cluster 6, see Table 2.4) is an aromatic L-amino acid decarboxylase necessary in the last step of serotonin biosynthesis (see section 1.3.1), also found in cricket heads Watanabe et al., 2011). Serotonin affects photosensitivity and seasonality in birds (Bhatt & Chaturvedi, 1992), photosensitivity in molluscs (Eskin & Maresh, 1982) and circadian rhythms in insects (Cymborowski, 1998; Page, 1987). Differential expression of genes in the serotonin synthesis pathway can be related to both, its direct role in the process and to its role as melatonin precursor. It is worth recalling that melatonin has been shown to have a role in photoperiodism at least in vertebrates and that different levels have been found in *A. pisum* under different photoperiodic conditions (Barberà et al., 2018).

Clusters 5, 7, 9 and 10 (see Table 2.4) are related to eye pig-

mentation and synthesis of opsins and retinal (see Figure 2.16), a key component of the visual pigment rhodopsin (see sections 1.3.1 and 3.3.1). NinaG, a glucose-methanol-choline oxidoreductase (cluster 5) is necessary for the chromophore synthesis of rhodopsin 1 in *D.melanogaster* (Sarfare et al., 2005; Ahmad et al., 2006), as well as PINTA[10], a protein with a CRAL-TRIO domain[11] (which is included in cluster 7, see Table 2.4) that is required for the biosynthesis of rhodopsin in *Drosophila melanogaster* (Figure 2.16) (Wang, 2005) and CYP 450 (Cytochrome P 450) (cluster 10) which is also involved in rhodopsin biosynthesis (see Figure 2.16). ABC transporters (cluster 9) are necessary for eye pigmentation (Ewart et al. 1994). As photosensitive proteins, opsins probably play some role in photoperiodism and circadian rhythms. Their expression might be influenced by light availability so as to fine tune sensitivity or to cope with it accordingly.

Cytochromes like CYP 450 (cluster 10, see Table 2.4) are not only related to rhodopsin biosynthesis (see Figure 2.16), but they are also well known for their role in detoxification and hormone biosynthesis (Schuler et al., 2007). They are involved in juvenile hormone (JH) (Sutherland et al., 1998; Feyereisen, 1999) and ecdysone synthesis (Schuler et al., 2007). JH is linked to aphid photoperiodic response

[10]Prolonged depolarization afterpotential is not apparent

[11]Cellular Retinaldehyde-binding Protein (CRALBP) and TRIO guanine exchange factor

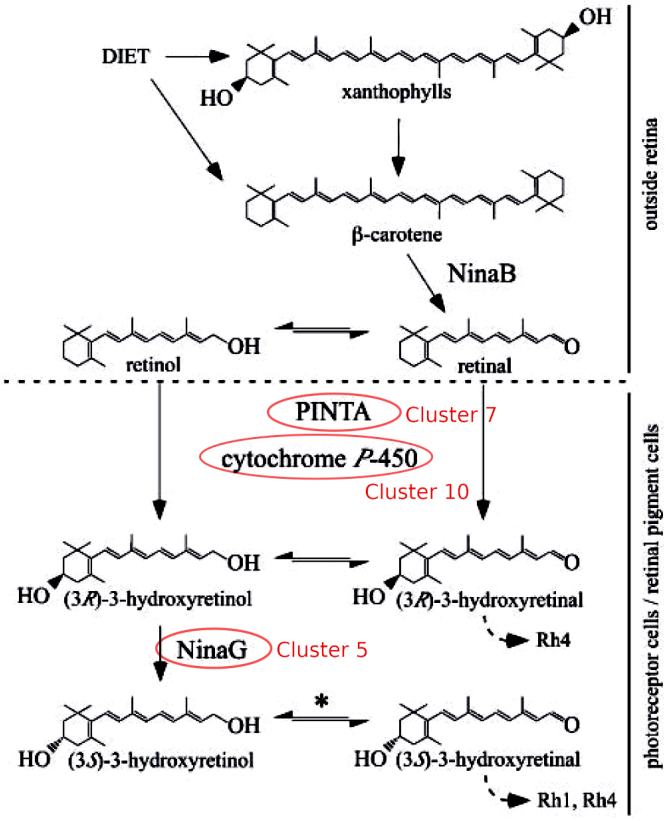


Figure 2.16: Biogenesis pathway of (3S)-3-hydroxyretinal in *Drosophila*. Genes belonging or related to clusters found in our study are circled. β-carotene in flies is obtained from the diet, but aphids can synthesize carotenoids (Moran & Jarvik, 2010). β-carotene is converted to retinal by NinaB outside of the retina. Retinal could interconvert with retinol. Both of them can be transported to the retinal pigment cells, where PINTA retinoid-binding protein, with a CRAL/TRIO domain (cluster 7) (Wang, 2005) acts in chromophore maturation. A cytochrome P450 monooxygenase-type enzyme (cluster 10) generates (3R)-3-hydroxyretinal from retinal or alternatively (3R)-3-hydroxyretinol from retinol. (3R)-3-hydroxyretinol is isomerized to (3S)-3-hydroxyretinol by NinaG, a glucose-methanol-choline oxidoreductase (cluster 5). (3S)-3-hydroxyretinal can serve as chromophore for two *D. melanogaster* rhodopsins: Rh1 and Rh4. Taken from Ahmad et al. (2006)

in aphids (Ogawa and Miura, 2014). In particular, the JHIII titer in *A.pisum* is lower in sexuparae[12] under SD conditions than in virginoparae under LD conditions (Ishikawa et al., 2011). Moreover, high levels of JH revert SD induction to parthenogenesis.

A wide variety of proteins with different biological functions (including neuronal differentiation) have Kelch domains (cluster 11) (Adams et al., 2000). Among the DE transcripts with Kelch domains, we found seven ACYPIs of gigaxonin type genes (Supplementary Tables S3 and S4). The gigaxonin type proteins in particular are related to neurodegeneration in giant axonal neuropathy (GAN) as they are necessary for proper axonal transport in mouse and human brain (Ding et al., 2006). The photoperiodic message in *A.pisum* is driven from the brain and must be transmitted to the developing embryos, likely along the axons of particular neurosecretory cells (Steel, 1977). Thus, it could be hypothesised that proteins with Kelch domains related to brain function might be participating in the movement of organelles and molecules through the nerve cells involved.

[12]Sexuparae: parthenogenetic females that give birth to the sexual individuals. See Introduction

2.5.2 Enriched pathways: KEGG tool

While DAVID uses KEGG together with other public genomic resources to generate biologically related transcript groups, it is useful to examine KEGG pathways results in particular (Tables 2.5, 2.6 and 2.7). This tool offers a pathway-centered context of the queried transcripts, facilitating a global view. Notwithstanding, many transcripts present in DAVID clustering tool are not included in KEGG pathways, as the latter is less comprehensive.

Three of the detected pathways had DE genes under both SD photoperiods: ABC transporters, peroxisome pathway and lysosome pathway (cystein-type peptidase activity)[13]. ABC transporters and lysosome related pathways were upregulated under LD conditions when compared to both SD conditions (i.e., they where downregulated under SD_{14} and SD_{10}). The peroxisome pathway, or more specifically, the pathway related to the cellular process of transport and catabolism involving peroxisome related genes, was found upregulated in LD vs SD_{14} but downregulated in LD vs SD_{10}. Although this result might seem contradictory, the genes upregulated in LD vs SD_{14} but downregulated in LD vs SD_{10} were not the same. Thus, this pathway could still be performing the same activity both under SD_{14} and SD_{10}.

[13]ABC transporters, peroxisomes and lysosomes correspond to DAVID clusters 9, 4 and 3 respectively

Consistently with our results found using the DAVID clustering tool (see sections 2.4.4 and 2.5.1), these pathways involving ABC transporters, peroxisomes and lysosomes (cystein-type peptidase activity) are affected by SD conditions (see Table 2.5). The genes involved in these pathways are downregulated when days are shorter. As seen in section 2.5.1, these enzymatic activities or proteins could have a role in brain remodelling, digestion, cuticle biosynthesis or eye pigmentation, all of them potentially relevant aspects of the photoperiodic response (as already discussed).

Some results indicate that LD photoperiod is a transcriptionally active state in the head, as revealed by increased transcription in pathways such as DNA replication, RNA transport, spliceosome and ribosome biogenesis (see Table 2.6). As mentioned in section 1.3.1, JH levels are high during long days, cause long day effects in aphids (Hardie et al., 1985; Ogawa et al., 2014) and promote ribosomal protein gene expression (Wang et al., 2017), which is consistent with our results, as we see that, at least, 17 genes from the ribosome biogenesis pathways in eukaryotes were found to be overexpressed in LD (see Table 2.6)

KEGG pathways analysis also revealed overexpression in LD of 6 genes from the FoxO (Forkhead transcription factors) signaling pathway (genes PRMT1, CycB, CDK2, SKP2, CycD2 and Plk4)(see Table 2.6). It is, however, difficult to determine if FoxO activity

is increased or reduced, as among those 6 genes there are FoxO inhibitors (CDK2, SKP2) as well as an activator (PRMT1). Moreover, of the other proteins found, one is inhibited (CycD2) and two are activated (CycB and Plk4) by FoxO activity. The scope of our study is insufficient to determine the outcome of the interaction between these elements and to confirm if FoxO itself is being repressed or activated. ACYPI numbers assigned to FoxO proteins[14] in the aphid were not found among our differentially expressed genes. We can, however, state that our results are in concordance with other findings that reveal FoxO transcription factors relevance in the photoperiodic response (Sim et al., 2015). FoxO acts downstream of insulin and JH (see Figure 1.19) in the response to photoperiod and diapause induction in mosquitoes and planthoppers (Sim et al., 2013; Sim et al., 2015; Yin et al., 2018). JH promotes the production of asexual morphs in aphids (see section 1.3.1), thus the upregulation of the FoxO signalling pathway under LD (see Table 2.6), where asexual parthenogesis is promoted, might be due to the effect of JH promoting inhibitory elements of the FoxO signaling pathway, as FoxO levels are low under LD conditions in other insects (see Figure 1.19).

Regarding the connection between FoxO and insulin in insects

[14] ACYPI008827, ACYPI005532, ACYPI008133, ACYPI006817 and ACYPI007895

(see Figure 1.19), it has been proposed that the insulin pathway is inhibited under SD conditions in aphids (Le Trionnaire et al., 2009). Also, two insulin-like peptide (ILP) coding genes were found to be underexpressed under SD conditions in *A. pisum* (Barberà et al., 2019). Moreover, ILPs activate JH production in mosquitoes (Sim & Denlinger, 2009), which as stated above inhibits FoxO.

We have found that DE genes are enriched in "insulin-like growth factor receptor signaling pathway" when comparing LD with both short photoperiods (see Tables 2.8 and 2.9). Also, using KEGG pathways we found genes related to the insulin resistance pathway to be downregulated under LD conditions (see Table 2.7). Again, although the outcome of the intricate interaction of these various elements of the insulin and ILPs pathway can't be elucidated with the information obtained in this study, it is remarkable to find the implication of these pathways in the photoperiodic response under both SD photoperiods taking into consideration that ILPs are the best known candidates to be the so called "virginoparin", the substance that triggers the production of asexual morphs, virginoparae, in aphids (Barberà et al., 2019).

It also result of great interest that FoxO has been proposed to be a general regulator in phenotypic plasticity mediated by insulin in *D. melanogaster* (Tang et al., 2011). Aphids, as stated in section 1.1, are well know for their polyphenism, a extreme case of

phenotypic plasticity, as one genotype can produce several different discrete phenotypes in response to environmental conditions. As in other animals, *D. melanogaster* reduces final body and organ size when developmental nutrition is diminished. This, however, doesn't occur in the genitalia where size is less affected because FoxO is inactivated. FoxO is part of the insulin pathway used to signal nutritional status (Tang et al., 2011). FoxO is, therefore, a crucial part of the mechanism that generates different phenotypes according to the environmental signal. It is possible, then, for this FoxO/insulin interaction to be mediating polyphenism in aphids (see section 1.1).

At least two genes related with the fly phototransduction pathway are DE, as seen also with the KEGG pathways results (see Table 2.5) and previously with the DAVID clustering tool (see section 2.5.1). Phototransduction is essential in the aphid seasonal response, as the main source of information is light. The light stimulus is detected through photosensitive proteins, such as cryptochromes and/or opsins (see section 1.3.1). In the case of opsins, their activation triggers a cascade that opens Transient Receptor Potential (TRP) and TRP-like (TRPL) channels to depolarize the cell membrane in photosensitive cells. It is suggestive that one of the two proteins related to the fly phototransduction pathway found by our KEGG pathways analysis (ACYPI008192) is an aphid TRPL channel, which

is overexpressed under LD conditions.

We also found transcripts related to the tyrosine pathways. Among these transcripts, tyrosine 3-monooxygenase (ACYPI008168) and aromatic-L-amino-acid decarboxylase (ACYPI009626) catalize the last two steps in dopamine synthesis (see General Introduction). Both are overexpressed in long photoperiod (LD). Dopamine is thought to be a transmitter in the photoperiodic induction of diapause (Wang et al., 2015), a messenger both in circadian and seasonal timing in insects (Danks, 2003) affecting, also, phototactic behaviour in aphids (Zhang et al., 2016). As mentioned in section 1.1, dopamine promotes photoperiodic response in some lepidopteran species. Thus, our results showing an increased expression of genes related to dopamine synthesis under LD conditions where aphids reproduce by parthenogenesis are consistent with the evidence found in other insect species. This evidence supports a model similar to that observed by Wang, et al. (2014) where dompamine and melatonin levels show and inverse relationship, as melatonin levels decrease under LD in aphids (Barberà et al., 2018)

Dopamine, is also involved in cuticle biosynthesis from tyrosine (Rabatel et al., 2013). It is thus possible for our findings to be also related with this process, as cuticule biosynthesis is one of the processes more frequently found to be affected by photoperiodism in aphids, as discussed above.

Three transcripts from genes involved in the tryptophan metabolism were found to be upregulated in LD when compared to SD_{14} (see Table 2.5). The tryptophan metabolism pathway includes an aromatic L-amino acid decarboxylase. As mentioned in section 2.5.1, this enzyme is necessary for serotonin and, eventually, melatonin biosynthesis, which are key metabolites in circadian rhythms and photoperiodism (at least in vertebrates, see section 1.2). As mentioned above, mutual inhibition between the melatonin and dopamine seems to regulate photoperiodic response in insects (Wang et al., 2014).

Three other pathways that can be associated with serotonin and melatonin are the "butanoate metabolism", "synthesis and degradation of ketone bodies" and "glycosphingolipid biosynthesis - ganglio series" pathways. Their connection, however, is less direct. They are related to acetyl-CoA, a cofactor needed in multiple biochemical reactions. Some of these reactions are relevant in the study of circadian rhythms and photoperiodism. Acetyl-CoA, for example, acts as a cofactor in an intermediate step of serotonin transformation into melatonin catalized by an insect arylalkylamine N-acetyltransferase (iAANAT) (Hickman et al., 1999). Butanoate metabolism and the synthesis and degradation of ketone bodies generate acetyl-CoA. The glycosphingolipid biosynthesis - ganglio series pathway is represented in our results by three genes which are

putative acetyl-coenzyme A transporters (see Tables 2.5 and 2.7). Thus, the three pathways are related to serotonin and metalonin through their involvement in acetyl-CoA production or transport.

The genes found in the ascorbate and alderate metabolism (see Table 2.7) catalize D-glucuronate synthesis. D-glucuronate is the substrate used for glucoronidation of diverse molecules. It has been proposed that glucuronidation of dopamine and serotonin plays a regulating role in their physiological function in brains in humans (Suominen et al., 2013; Ouzzine et al., 2014). Moreover, among the 23 genes DE in the "metabolic pathways" category (see Table 2.7), three UDP-glucuronosyltransferases (which catalyze the glucuronidation reactions) were found[15].

Several pathways and genes related to the serotonin and melatonin pathways were identified by the KEGG pathways analysis. Some of these pathways, were upregulated in LD, as the synthesis and degradation of ketone bodies, while others where downregulated in LD. The outcome in the interaction between these pathways to produce high or low levels of serotonin and/or melatonin goes beyond the possibilities of this study, given the complexity of the interaction and the indirect role of some elements in the process. It is, however, possible for us to confirm an effect of photoperiod in the pathways involving serotonin and melatonin, which are key

[15] ACYPI001041, ACYPI002054 and ACYPI009222

molecules in the study of the photoperiodic response.

2.5.3 Enriched GO terms involved in photoperiodism

Using REVIGO analysis tool, we were able to identify several Gene Ontology (GO) terms which were enriched under different photoperiodic conditions (see Tables 2.8 and 2.9). As the DAVID tool failed to include all ACYPIs in the analysis, we extracted all GO terms associated to differentially expressed ACYPIs to be the input in REVIGO. At least 9 and 12 GO terms (for SD_{14} and SD_{10} respectively) are directly related with light sensitivity and eyes. Six of them are shared among both short photoperiods: visual behavior, eye pigment granule organization, R8 cell development, R7 cell development, ommatidial rotation and eye-antennal disc morphogenesis. Other GO terms are involved in hormone synthesis and brain function, as has been also found using the DAVID clustering tool (Table 2.4) and the KEGG pathways tool (Tables 2.5 - 2.7). When comparing both short photoperiods with LD, REVIGO found that the DE expressed genes are enriched in GO terms related with the insulin-like growth factor receptor signaling pathway (see Tables 2.8 and 2.9).

Most interestingly, in both SD conditions, which produce sexual females and males, REVIGO found enrichment of genes related to

mating behaviour (GO:0007617) (see Tables 2.8 and 2.9) and in the male producing photoperiod specifically (SD_{14}) male courtship behaviour related GO terms were found (GO:0008049) (see Table 2.8).

2.5.4 Top differentially expressed genes

Although KEGG pathways, DAVID and REVIGO tools allow for a broader view of differentially expressed groups of genes, we considered interesting to analyse certain individual transcripts because of their different relative level of expression in different photoperiods. Tables 2.2 and 2.3 showed the top under and overexpressed genes sorted by their $\log_2$ fold change in expression in two SD conditions compared with LD. These lists also include transcripts without $\log_2$ fold change value that were only found in one of the conditions. This indicates that they are only expressed under a particular photoperiod. Thus, when calculating the fold-change value, the result is "infinite" as there is a number divided by zero. These results are expressed as inf. or -inf. in the Tables 2.2 and 2.3, depending on the condition where the transcript is absent. Hypothetical connections of the top differentially expressed genes with the process of photoperiodic induction in aphids is discussed below.

The *takeout* (*to*) gene is differentially expressed in male-producing sexuparae

Interestingly, among the top 10 transcripts with higher expression level differences shown in Table 2.2, there are three *takeout*-like genes (ACYPI060748, ACYPI001035 and ACYPI003395) that are underexpressed in the male inducing photoperiod (SD_{14}) (see Table 2.2). Those three genes showed a 8.24; 8.05 and 6.47-fold decrease respectively under SD_{14} conditions when compared to LD. Moreover, not only are these genes underexpressed in the male inducing condition, two of them, ACYPI001035 and ACYPI00339, are also overexpressed in the sexual female inducing photoperiod SD_{10} (although they are not found among the top DE genes) (see Supplementary Table S2).

D. melanogaster's *takeout* (*to*) is an output gene of the circadian clock, a target of the central clock transcription factors (Sarov-Blat et al., 2000; So et al., 2000). In flies, it is regulated by CLK (So et al., 2000; Benito et al., 2010) and by *Pdp1* (Benito et al., 2010), both being core clock proteins. *D. melanogaster*'s *to* has an E-box bound by CLK-CYC and its mRNA and protein levels decrease in mutants lacking CLK and CYC activity. *Pdp1* expression increases together with *to* transcript levels in flies (Benito et al., 2010). Previous studies in our group found that *Pdp1* gene levels of expression

increase under short day conditions (12L:12D or SD_{12}) that promote sexual females production in *A. pisum* (Barberà et al., 2017). As mentioned above, the sexual female inducing photoperiod used in this study, SD_{10}, shows an increase in expression of two *to*-like genes (ACYPI001035 and ACYPI00339), in accordance with *Pdp1* and *to* expression levels increasing and decreasing simultaneously, probably due to *Pdp1* mediated regulation.

The TAKEOUT amino acid sequence shows similarity to haemolymph juvenile hormone binding protein JHBP (Sarov-Blat et al., 2000) and it is thought that JH transport and/or nuclear action could be regulated by the JHBPs availability (Steel & Vafopoulou, 2002). Therefore, it is possible that *takeout* could be affecting the photoperiodic response through its connection with the circadian clock and/or via JH regulation.

Sarov-Blat et al. (2000) also found that *to* provides nutritional status information in flies. In this context, due to its relation with the circadian clock, the authors propose that this gene could be contributing to the anticipation of food availability. Although the photoperiodic response is mainly triggered by the day-night cycle, it is affected (to a lesser extent) by temperature. In a similar manner, nutritional status affects photoperiod response in other insects (Reznik et al., 2015), but it does not affect the circadian clock (R. Nelson et al., 2010). It is possible, then, that *to*-mediated

food availability information is also used and weighed by the photoperiodic response mechanism in aphid seasonality. Not only is *to* related to clock and photoperiodic-response genes, but also at least three members of the *takeout* gene family are involved in sex-specific processes in *D. melanogaster* heads, and not in other tissues (Dauwalder, 2002). Moreover, *to* is regulated by *doublesex* (*dsx*) in flies (Dauwalder, 2002), which is also responsible for sex-specific traits in other insects like beetles (Kijimoto et al., 2012) and, interestingly, in the photoperiod-driven sex determination process in another invertebrate with cyclic parthenogenesis: *Daphnia pulex* (LeBlanc & Medlock, 2015). As previously mentioned (see section 2.1), SD_{14} photoperiod induces male production. It is then possible that this photoperiodically-induced sex determination might be driven or affected by one or more members of the *takeout* gene family in aphids, after regulation by *dsx*. To our knowledge, this is the first study to detect an effect of photoperiod on the *to* gene in aphids. A very recent study in *Locusta migratoria* also found DE of *takeout* in the central nervous system and proposed its involvement in the inhibition of maternal diapause induction of *L. migratoria* under SD photoperiods (Jarwar et al., 2019).

Acyrthosiphon pisum LIN-28 homolog

The aphid protein corresponding to ACYPI009849 (downregulated in SD_{14}, see Table 2.2) is a homolog of LIN-28, an RNA-binding protein known to interact with a microRNA named *lethal-7* (*let-7*) (Tsialikas & Romer-Seibert, 2015). LIN-28 binds *let-7* pre-microRNA, thus interfering with mature *let-7* microRNA production (Piskounova et al., 2011, Stratoulias et al., 2014). This microRNA mediates timing of larva to adult transition in *D. melanogaster*, and has been proposed to be "part of an ancestral pathway controlling the transition from larval-to-reproductive animal forms" (Chawla & Sokol, 2012). If this is the case in aphids, it might play a relevant role in diapause induction or induction of sexual forms. Moreover, *let-7* regulates the circadian rhythm by repressing the clock related gene named *clockwork orange* (*cwo*) in *D. melanogaster* heads (Chen et al., 2014). In turn, *let-7* levels of expression are regulated via prothoracicotropic hormone (PTTH), a possible effector in the aphid photoperoidic clock output mechanism (Chen et al., 2014). It would be worth studying a putative role of *lin-28* gene and its influence in let-7, given its possible connection with both developmental transitions and circadian rhythms in aphids.

Uncharacterized transcripts are involved in photoperiodic response

In our ranking of differentially expressed genes many correspond to uncharacterized genes (see Tables 2.2 and 2.3).

Under SD_{14} photoperiod, 10 transcripts with no BLAST hits or known conserved domain were found, and 15 under SD_{10} conditions. Given that the molecular mechanisms ruling the photoperiodic response are not well understood, it is not surprising that many of the elements involved are not characterized. Although a high $\log_2$ fold change does not necessarily entail relevance in the process, it does provide good candidates to search for elements involved in the photoperiodic response. Further research on these candidates might provide us with key molecules in the process.

Of the top DE genes, only one of the entries is shared in both SD conditions: ACYPI072820, but no hits were found using BLAST nor in conserved domain (CD) searches. Further specific research on this transcript could reveal how relevant it is, given that it is overexpressed both in male (SD_{14}) and female producing sexuparae (SD_{10}).

There are several difficulties while studying the function of orphan genes or unknown proteins. Further experiments that analyse the expression of these unknown genes under controlled systems

could shed some light on their role in aphid photoperiodism. Insect cell cultures can provide useful tools to asses the interaction of these proteins with other known molecular elements involved in biological rhythms. Also, silencing target genes (e.g. via RNA interference, RNAi) is a valuable tool in studying the effect of loss of gene function that has already been tested in aphids (Jaubert-Possamai et al., 2007) but needs to be fully developed.

Troponin C

It is puzzling to some extent the finding of differential expression of a gene coding for troponin C (ACYPI009055). It was found to be overexpressed in SD_{10} (see Table 2.3). Troponin C function is typically related to muscular activity. There are, however, a few studies that refer to its presence in brain (Fine et al., 1975; Lyckman et al., 2008 and Johnston et al., 2017). Ocular dominance (the preference of one eye over the other to provide visual input) can be modified by covering one eye during a critical period of postnatal development in mammals. Lyckman et al. (2008) found that, in mice, Troponin C is differentially expressed in the primary visual cortex during postnatal critical period for ocular dominance plasticity, and refer to it as a "novel neuronal protein". Its expression is rapidly reduced in response to brief visual deprivation. It is, thus, involved in the effect of the input of light on the brain, more specifically in

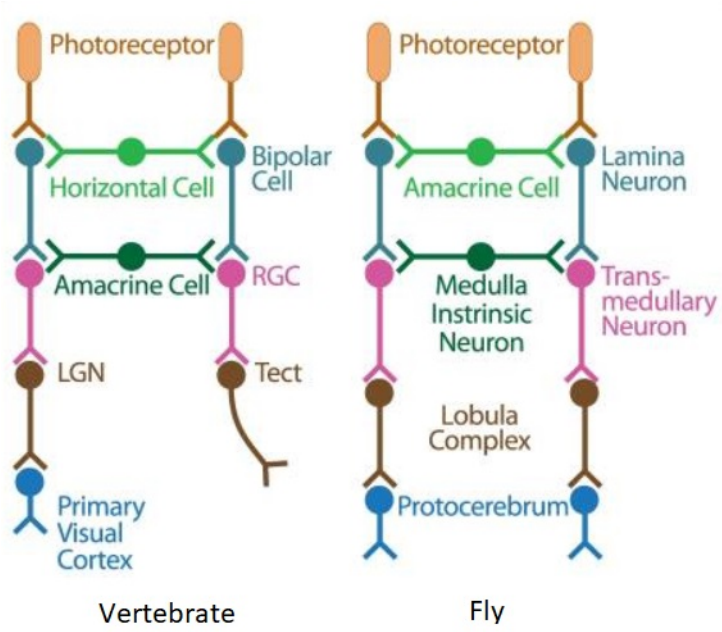


Figure 2.17: Similarities between vertebrate and *Drosophila* transduction of visual information to the brain. The light signal from the photoreceptor reaches the primary visual cortex and the protocerebrum. RGC, Retinal ganglion cell; LGN, lateral geniculate nucleus. Adapted from Sanes & Zipursky (2010).

the primary visual cortex.

Among the similarities between vertebrate and insect visual systems, both in mammals and flies, the light received by photoreceptors is transduced into electrical signals that eventually reach the visual cortex (in mammals) or the protocerebrum in flies (Sanes & Zipursky, 2010) (Figure 2.17). Protocerebrum is where the aphid photoperiodic receptor has been proposed to be located (see section 1.4). It is, thus, interesting for us to have found differential

expression of troponin C in the aphid protocerebrum, an analog structure of the mammal visual cortex, where troponic C was also found to be regulated by light exposure. Sanes & Zipursky (2010) propose that troponin C promotes synaptic stability in mammals. However, any parallelism between insects and mammals should be taken cautiously for many reasons: both structures (mammal visual cortex and insect protocerebrum) don't share exactly the same developmental origin, the visual system is apparently not involved in the aphid photoperiodic response and overexpression found in mammals was not related to the photoperiodic response (Lyckman et al., 2008).

Brain development

Several transcripts found in this study are involved in brain modelling and development. Troponin C (ACYPI009055), mentioned above, is used in mammal brains to define ocular dominance. Peroxidasin (ACYPI21941) and the proteins coded by ACYPI48450 and ACYPI21942 are upregulated under SD_{10} photoperiod. The three of them have leucine-rich repeats and immunoglobulin-like domains, which are common to peroxidasins (Nelson et al., 1994; Soudi et al., 2012). Peroxidasin specific function is still elusive, but it plays a role in brain development in *D. melanogaster* (Soudi et al., 2012), allowing correct ventral nerve cord condensation. The protein

coded by ACYPI21942 could also be involved in gene regulation (see section 2.5.4).

In relation to brain development and modelling, where cellular division is a key process, three ACYPIs involved in the cell cycle were found to be differentially expressed in LD vs SD_{14}: ACYPI34702, ACYPI004418 and ACYPI27105. The first protein (ACYPI34702) is the N-acetyltransferase ESCO2. It is known as *eco* in yeast (Vega et al., 2005) and *deco* in the fly *D. melanogaster* (Williams et al., 2003). It is requiered for the establishment of cohesion of sister cromatids during DNA replication. Its underexpression in SD_{14} might reflect a decrease of cell division. The other two ACYPIs (ACYPI004418 and ACYPI27105) have conserved domains related to cell division and the cell cycle (FtsK/ SpoIIIE and ZipA superfamily respectively).

Gene regulation and chromatin structure

Many genes seem to be directly associated to gene regulation via diverse mechanisms, including protein and nucleic acid degradation or chromatin modification.

Two histones, H1 and H2A (ACYPI006931 and ACYPI008799) were found to be overexepressed under LD photoperiod (or underexpressed in SD_{10}). Proper DNA packaging is essential and dependent on histones, whose synthesis must be well balanced for proper chromosome replication (Rattray & Müller, 2012). Histone transcription

is induced in the S-phase of the cell cycle. Thus, an important difference in histone transcription seems to indicate an increase in cellular division. This increase in cell division might be due to a close moulting process affecting external structures. It has also been proposed that gene expression might be actively regulated through histone availability in certain physiological processes (Prado et al., 2017). Whether the photoperiodic response is regulated to a certain extent via histone availability is unknown.

Also related to histones, ACYPI30700, the Craniofacial development protein 2 (cfdp2) was found to be overexpressed in SD_{10} photoperiod. It belongs to the Bucentaur familiy (BCNT), which is essential in chromatin organization and function (Messina et al., 2014a). YETI, a member of BCNT in *D. melanogaster* affects the accumulation of the histone H2A (Messina et al., 2014b).

Ubiquitins are well known for their role in protein degradation (Hershko & Ciechanover, 1998). In such a metabolically active process as the photoperiodic response can be, it seems natural to find ubiquitins such as ACYPI47959 to be differentially expressed (see Table 2.3). They target proteins for degradation, thus being actively regulating their action. The ACYPI007067 contains a domain belonging to the CTD superfamily, for pre-mRNA processing, which is also actively involved in gene regulation.

The protein coded by ACYPI007433 (underexpressed in SD_{14},

see Table 2.2) is known as "Lupus La protein". As the name suggests, it has been studied for its relation with lupus disease, but also with Sjogren's syndrome (Teplova et al., 2006). This protein is poorly studied in insects, but is also known for its role in the metabolism of RNA and could thus probably be regulating gene expression in the photoperiodic response at the mRNA level.

Also related to nucleic acid metabolism, ACYPI002693 is a deoxycytidylate deaminase, involved in pirimidine nucleotide metabolic process.

ACYPI004446 is a chromobox protein homolog 1 (CBX1), which belongs to the Heterochromatin Protein 1 (HP1) family (Lomberk et al., 2006). This family of proteins interacts with chromatin for transcriptional regulation, chromatin modification and replication to DNA repair (Lomberk et al., 2006).

The transcript identified as ACYPI21942 was mentioned previously due to its putative role in brain development because of its leucine-rich repeats and immunoglobulin-like domains. The protein encoded by this transcript also contains a GATA zinc finger domain, which is a protein recognition motif often found in transcription factors (Matthews & Sunde, 2002). It is probably involved in gene regulation.

Cuticle related proteins

Cuticle synthesis in insects involve daily deposition of chitin layers. Studies in aphids have also found that seasonal photoperiodism affects cuticular proteins (Le Trionnaire et al., 2007; Cortés, 2010). Transcripts ACYPI006355 and ACYPI007152 code for proteins related not only to gene regulation and cromatin structure (as mentioned in the previous section), but also to cuticles (cuticiclin and chitin binding superfamily respectively). Both could be involved in the daily deposition of cuticle layers and be affected by photoperiodism.

In relation with cuticle biosynthesis and biological rhythms, it is known that in many insect species, the cuticle is thickened every day by depositing a pair of chitin layers, each with a different orientation (Neville, 1983). Cuticle deposition rhythm in the bean bug *Riptortus pedestris* and in *Drosophila melanogaster* is regulated by a circadian clock. In the study by Le Trionnaire et al. (2007), the majority of the significantly regulated transcripts found to be differentially expressed belonged to cuticular proteins. Moreover, Cortés et al. (2008) showed highly significant overexpression of cuticular proteins under SD photoperiods in aphids.

General functions

Many of the results analysed consist of proteins found in a wide variety of biological processes involving numerous substrates, thus specific functions or candidate roles are difficult to ascertain.

Leucine-rich proteins, such as ACYPI002751 (overexpressed under SD_{10} and underexpressed under SD_{14}, see Tables 2.2 and 2.3), are involved in a wide variety of biological functions (Kobe & Deisenhofer, 1994). They are often involved in protein-protein interactions and in signal transduction pathways (Kobe & Deisenhofer, 1994), which are necessary in a complex process as the photoperiodic response. Interestingly, among the top results of a BLAST search of ACYPI002751 against the *Drosophila* genus we found three acid labile subunits (ALS) of the insulin-like growth factor-binding protein complex. *Drosophila* ALS is required for the growth and metabolic functions of the insulin -like peptides in the fly (Arquier et al., 2008). As mentioned above (section 1.3.1) the insulin pathway and ILPs are of great relevance in photoperiodic response.

The transcript associated to ACYPI009171 (see Table 2.2) codes for the Embryonic development factor, which is a homolog of Use1 (Unconventional SNARE in the Endoplasmic reticulum 1) (Dilcher, 2003), a vesicle transport protein. There is little information available on this gene in insects. It is probably involved in diverse

biological functions and a variation in transcript levels most likely reflects changes in general cellular activity (Dilcher, 2003).

The transcript ACYPI060277 (see Table 2.2) has a CYTH-like domain, also known as triphosphate tunnel metalloenzyme (TTM)-like phosphatase. The CYTH-like superfamilly hydrolizes a wide variety of substrate containing triphospates, thus it could be involved in several processes.

Digestive and detoxification activity

Head preparations include salivary glands and mouthparts. When they feed, aphids elicit a defense response in plants and perform the first steps of the digestive process in mouthparts, thus they make use of detoxification and digestive enzymes to help in extra-oral digestion of phloem sap (Carolan et al., 2011; Zhang et al., 2017). Some of these proteins were found in our analysis. They might be affected by photoperiod indirectly, as food quality or feeding behaviour may vary when plants are exposed to shorter photoperiods. Although the effect of the photoperiod on plants has been proven not to affect the photoperiodic induction of sexual morphs of the aphids feeding on those plants (Beer et al., 2017), it could be affecting other processes, such as the digestive enzyme production.

Also, it is possible for these enzymes to perform different roles related to photoperiodic response, as some have been found to be

expressed in tissues or organs such as the brain, which are not in direct contact with food or plant toxins.

The protein coded by ACYPI000002 (see Table 2.2) is a sucrase thought to be the main enzyme used to digest sucrose in *A. pisum* guts (Price et al., 2007). It is problably playing its role as digestive enzymes in the salivary glands and saliva.

Another transcript, ACYPI010188 (see Table 2.2), belongs to the family of glucose dehydrogenases, which were also found in *A. pisum* saliva (Carolan et al., 2011).

As mentioned in section 2.5.1, cytochromes (like CYP 450) are related to detoxification and hormone biosynthesis. Thus, its overexpression found at SD_{10} (ACYPI000949, see Table 2.3) might be related with hormonal control of the photoperiodic process. It is, however, more likely for us to have found CYP 450 to be DE due to its role in detoxification of plant defense proteins. Previous studies analyzing composition of the aphid salivary glands have already found CYP 450 in *A. pisum* and other aphids (Carolan et al., 2011; Zhang et al., 2017).

Maltase (ACYPI007753, see Table 2.3) is a digestive enzyme affected by the circadian clock in flies (Ciuk et al., 2009). It could therefore be affected by such a strong difference in day length when comparing LD to SD_{10}. Digestive enzymes might be detected for their presence in salivary glands.

ACYPI010224 (see Table 2.2), with two peroxidase domains, is probably used for detoxification of reactive oxygen species (ROS) generated by the plant (Zhang et al., 2017). A venom protease, ACYPI006282 (see Table 2.2) is also probably related to detoxification.

Serine proteases (ACYPI000220, ACYPI004807) are among the top differentially expressed genes. Similar proteases have been found in aphid salivary glands (Carolan et al., 2011; Zhang et al., 2017). As discussed in section 2.5.1, serine proteases might also be involved in processing neuropeptide or peptide hormones. They are overexpressed in SD photoperiods and their action could be part of the process that leads to sexual morph production. ACYPI003916 (serin protease 4 trypsin down in SD_{10}) has also been found in salivary glands (Carolan et al., 2011; Zhang et al., 2017).

Cathepsins, ACYPI000021 and ACYPI001175 were under and overexpressed in the shortest photoperiod (SD_{10}), respectively. Cathepsins are involved in many physiological aspects, including stress response, development and digestion and some are regulated by ecdysteroid hormones (Saikhedkar et al., 2015; Ge et al., 2014). Thus, the expression of one cathepsin could be activated while another one repressed under the same stimulus, such as shortening of the photoperiod.

Opsins and cryptochromes

Although no significant result was found (except for one particular type of aphid opsin named Ap-arthropsin, see section 3), there seems to be a clear trend in opsins and cryptochromes detected in the transcriptomic analysis (Figure 2.15). These light sensing proteins seem to have their transcription increased in coordination with the shortening of the photoperiod. The shorter the photoperiod, the higher the expression of opsins and cryptochromes. If this trend is confirmed by other experiments, it would suggest that aphids make more use of photosensitive proteins as winter approaches. This might increase the sensitivity of the insect if protein availability implies better detection. The differences observed, however, are slight. Opsins and cryptochromes seem to show low expression levels, thus making their analysis more difficult and hindering the detection of significant differences.

2.5.5 Molecular pathways leading to sexual females or males production

Shortening of the photoperiod produces the two different sexual aphid morphs: males and sexual females (oviparae). The ratio at which they are produced, however, is different according to how short that photoperiod is. Although the signal to produce one or the

other is the same (shortening of the photoperiod), the quantitative value of that signal is relevant for the outcome. A slight shortening (SD_{14}) produces only males while a more considerable shortening (SD_{10}) produces sexual females. An intermediate short photoperiod of 12L:12D (SD_{12}) is able to produce both males and sexual females (see Figure 1.3). Are these two different outcomes (males or sexual females) produced by the same molecular pathway differing only in some crucial steps? or are males and females produced from entirely different independent pathways? In support of the possibility for a common mechanism, our study reveals many shared pathways between the two SD conditions. Small differences in the molecular pathway triggered by the cumulative effect of light (or darkness) could lead to very different outcomes. Specific genes, such as the *takeout*-like found in the male-producing photoperiod might be one of this fundamental differences. On the other hand, the number of DE genes in SD_{10} triples that of SD_{14} (1072 vs 329). As mentioned earlier, this is most probably caused by SD_{10} being much more different to the reference photoperiod, LD. How many of all those extra genes are necessary to generate a sexual female progeny? If the whole pathway needed to produce females is completely independent to that of males, then those extra DE genes could contain the elements forming that pathway. We consider, however, the latter scenario to be less likely, as most of the gene clusters found in SD_{14}

are also present in SD_{10}, which means that most of what is needed to produce males is also present under SD_{10} conditions. In fact, as seen in Figure 1.3, although the *A. pisum* strain YR2 under SD_{14} and SD_{10} produces mainly males and sexual females respectively, both photoperiods are capable of producing the sexual morphs, albeit one of them in lower proportion. Therefore the pathway is most probably shared, with minor differences that result in different sex ratio.

2.6 Main conclusions of the transcriptomic analysis

Previous transcriptomic studies found several genes and pathways related to the aphid photoperiodic response (Cortés et al.,2008; Le Trionnaire et al., 2009; 2012; Ji et al., 2016). Apart from housekeeping functions (cytoskeleton and ribosomal protein-related genes) those studies found more specific and informative terms, such as those related to cuticular proteins, the neuroendocrine system, hormones (dopamine and JH), photoreception and food processing. Our study supports some of these findings and expands the list of specific genes involved in the photoperiodic response.

When differential gene expression was analyzed using different clustering tools to detect biological processes involved in photoperiodic induction of sexual reproduction in aphids, we found that aphids exposed to two different short photoperiods differentially express genes related to hormone, neuropeptides and neurotransmitter biosynthesis, some of them important candidates to be involved in the process, such as serotonin, JH, dopamine and melatonin. Moreover, consistently with previous findings by our group (Barberà et al., 2019) we found several genes involved in the insulin pathway or insulin-like peptides (ILPs), which is in accordance with the possible role of ILPs in virginoparae production. Consistently with

a response induced by light availability, we also found enrichment of pathways responsible for light perception, such as those needed for opsin synthesis and phototransduction. It is also evident from our results that photoperiod has an impact on brain functioning, as revealed by the enrichment of pathways related to brain remodelling and increased transcription and gene regulation. Digestive enzymes were found to be affected by the length of the photoperiod, which might be related to a putative effect of the light regime on the plants where aphids were fed. With the detailed analysis of top DE genes we found genes likely involved in the aforementioned processes. Also, some specific transcripts drew our attention, as the three *takeout*-like genes underexpressed in SD_{14} that could be related to male progeny production. Troponin C is also worth mentioning, as previous studies found it to be regulated by light availability in the visual cortex of mice and we found it to be overexpressed under SD_{10} in an analogous structure, the aphid protocerebrum. Further analysis is needed to understand the specific role of many uncharacterized transcripts (>40%) found to be among the top DE genes in the photoperiodic induction of sexual aphid morphs.

Chapter 3

Candidate genes in photoperiodic photosensitivity: opsins and cryptochromes

3.1 Introduction

In order to detect shortening of the photoperiod and anticipate to the harsh winter season, aphids must be able to sense light and day length. This input signal must be processed and somehow transduced to a chemical message that is then interpreted by target structures to perform the necessary tasks to cope with winter (i.e. change the developmental pathway of the embryos to develop as sexual morphs). Light detection in the photoperiodic process requires photosensitive molecules that have been located in the aphid protocerebrum rather than in the eyes (see section 1.4). The identity of this molecule or whether there is more than one type of molecule involved remains unknown. Two families of protein photoreceptors seem good candidates to be involved in photoperiodic photosensitivity: opsins and cryptochromes (Terakita, 2005; Michael et al., 2017). These proteins could be involved not only in the photoperiodic response, but also in circadian rhythms, as is the case in other organisms. Understanding their possible involvement in photoperiodism and circadian rhythms in aphids requires characterization of their gene structure and expression patterns. Fortunately, not only are aphids photoperiodically sensitive, but they also they have anholocyclic lineages that are unable to respond to photoperiod shortening, thus reproducing continuously through

parthenogenesis. These lineages are natural mutants of great value to study the photoperiodic response. By studying what prevents an anholocyclic lineage from producing sexual morphs, we can obtain hints that help us understand the cascade of events leading to the production of winter morphs. One possibility is that differences at the molecular level between holocyclic and anholocyclic lineages that ultimately trigger a different response to the same photoperiod might be found in the sensitive phase of the process, thus rendering these anholocyclic strains "blind" to the shortening of the photoperiod. On the other hand, they might be able to detect shortening of the photoperiod but unable to respond, meaning that the photoperiodic response is stopped at later stages in the induction process in these strains.

3.2 Objectives

The general objective of this chapter is to identify and characterize the repertoire of opsin and cryptochrome genes in aphids and study their expression in relation to photoperiod. These are the main photosensitive protein families generally thought to be related to photoperiodism and circadian rhythms in insects (Terakita, et. al, 2005; Saint-Charles et al., 2016; Ni et al., 2017; Michael et.al, 2017). We also aim to identify their seasonal, daily and anatomical patterns of expression to assess their possible relation with the photoperiodic response. We have established the following specific objectives:

1. Identify and validate genes encoding opsins and cryptochromes in the genome of *A. pisum*.

2. Quantify and compare opsin and cryptochrome gene expression at different times of the day in holocyclic and anholocyclic strains under different photoperiodic conditions to study their possible connection with circadian clocks and photoperiodism.

3. Localize the site of expression of opsin and cryptochrome genes to asses and study their relation with anatomical structures involved in the photoperiodic response.

In this chapter, we describe in depth the development of these objectives. In first place (section 3.3), we present the work carried

out for the opsin gene family. In the second half of this chapter (section 3.4), we focus on the cryptochrome genes of the pea aphid.

3.3 Opsins

3.3.1 Introduction to opsins

Light sensors or photoreceptors are key components of the input pathways in both circadian and photoperiodic systems since they are essential to keep both rhythms synchronized with the environment (see Figure 1.12). Dietary-deficient experiments suggest that vitamin A-based pigments convey photoperiodic information into the insect photoperiodic system (Saunders, 2002; Nelson et al., 2010). The first experiments on the southwestern corn borer *Diatraea grandiosella* found that diapause induction was affected when fed a carotenoid-free diet (Takeda, 1978). Other insects were later shown to behave similarly, needing carotenoids to induce a photoperiodic response (Saunders, 2002). Most interestingly, aphids are among the very few animals that synthesize their own carotenoids (Moran et al., 2010), which are needed for vitamin A synthesis (see Figure 2.16). Opsins are well characterized photoreceptive proteins that use retinal (a vitamin A derivative) as a chromophore. This makes opsins strong candidates to provide the photoperiodic system with day length information (Saunders, 2002; Nelson et al., 2010). Opsins belong to the G-Protein coupled receptor (GPCR) superfamily (Attwood et al., 1994) comprised of seven-transmembrane domain proteins

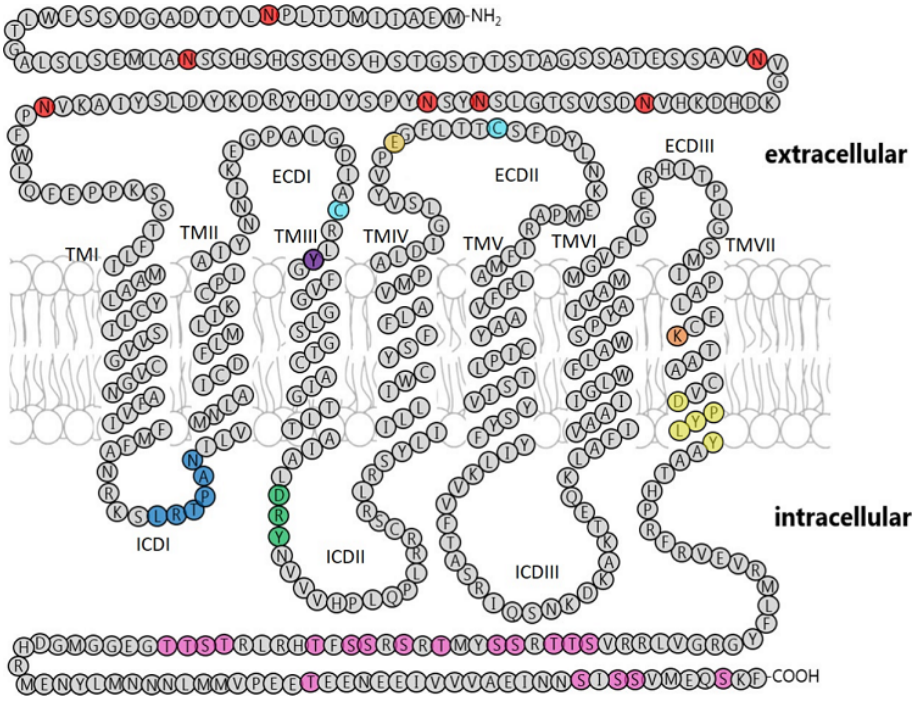


Figure 3.1: The tertiary structure of RH7 from *D. melanogaster* shows most of the conserved features in opsins: The 7 transmembrane (TM1 to TM7) domains typical of G protein–coupled receptors (GPCRs); in orange, the conserved Lysine in TM7 domain typically found in opsins (Gärtner and Towner, 1995); Counter-ion E181 in the second extracellular domain (ECDII) (Shichida et al., 2013); the cysteins in the extracellular domains I and II (ECDI, ECDII) (Gärtner and Towner, 1995; Hwa et al., 1999); the D/ERY motif in the intracellular domain II (ICDI) (Franke et al., 1990); and the N/DPxxY in the seventh transmembrane helix (Mirzadegan et al., 2003). Modified from Senthilan and Helfrich-Förster (2016).

(Figure 3.1). Within the GPCR superfamily, opsins constitute a monophyletic group characterized by the presence of a conserved lysine in the seventh transmembrane helix to which retinal is bound (see Figure 3.1).

Duplication events followed by sequence divergence have generated a complex diversity of opsin types through animal evolution (Porter et al., 2011; Terakita et al., 2011). Phylogenetically, opsins can be divided into four major lineages (Porter et al., 2011): the "C-opsins" (described from ciliary photoreceptors), the "R-opsins" (rhabdomeric), the cnidarian opsins ("Cnidops"), and a mixed "Group 4 opsins" (Gr4). The C and R-type visual opsins were the first to be characterized (Porter et al., 2011), and first thought to belong to vertebrates and invertebrates, respectively. Further information revealed that none of them is exclusive to either vertebrates or invertebrates, and that they also include non-visual opsins (Porter et al., 2011; Terakita et al., 2011). C-type opsins include the vertebrate visual pigments, which are expressed in ciliary photoreceptors. These pigments form sub-clusters according to their peak of light wavelength absorption: short (ultra-violet), medium (blue) or long (green-yellow) wavelength sensitive opsins. C-opsins have also been found in other cell types and in some invertebrates such as arthropods, molluscs, echinoderms and hemichordates (Porter et al., 2011). R-type opsins are typically found in rhabdomeric photoreceptors, like arthropod and cephalopod visual pigments. However, some rhabdomeric opsins have also been found in vertebrates, like *Xenopus laevis* and humans (Provencio et al., 1998; 2000). Group 4 opsins (Gr4) is a functionally diverse group consisting of 'reti-

nal G-protein coupled receptors' (RGR), peropsins and neuropsins, with representatives from chordates, molluscs, echinoderms, and hemichordates (Porter et al., 2011). Cnidopsins is the only group that still lacks representation on taxonomic categories other than cnidarians and ctenophores (Plachetzki et al., 2007). Most opsin studies have been carried out on preparations from retinal tissues, thus information on their non-visual counterparts is scarce (Porter et al., 2011). Moreover, there has been a lack of diversity in the organisms used as subject of study. In hexapods, holometabolous insects such as dipterans (Posnien et al., 2012), hymenopterans (Wang B. et al., 2013) and lepidopterans (Briscoe, 2001; Everett et al., 2012) have received more attention. Fewer studies have focused on hemimetabolous insects like the cricket *Gryllus bimaculatus* (Henze et al., 2012) and *Modicogryllus siamensis* (Tamaki et al., 2013) or the aphids *Megoura viciae* (Gao et al., 2000) and *Acyrthosiphon pisum* (Döring et al., 2011; Li et al., 2013).

To investigate the possible involvement of opsin-based photoreceptors in aphid photoperiodism we identified the full aphid opsin gene repertoire in the *A. pisum* genome, experimentally characterized their coding sequences and quantified their expression in aphid heads at different times along the day-night cycle, under different photoperiodic regimes and in two different aphid strains, one holocyclic and one anholocyclic. Additionally, we attempted to

amplify opsin transcripts through PCR in dissections from aphid heads, antennas, central nervous system and embryos as an indirect way of knowing their site of expression. To show a more specific localization, we performed *in situ* hybridizations of six aphid opsin genes on the central nervous system of the pea aphid.

3.3.2 Materials and methods

Identification of opsin genes present in the pea aphid genome

Aiming at identifying the whole opsin repertoire present in the aphid *A. pisum*, protein sequences representative of the different opsin types from different organisms (Table 3.1) were used in BLASTP searches against the *A. pisum* Reference Sequence (RefSeq) collection of peptides available at the Aphidbase database[1]

The identified sequences were used to search for similar sequences in the genome of the aphid *Myzus persicae*[2]. Aphid sequences were aligned with a set of protein sequences including representative of both visual and non-visual opsins from different organisms. Protein sequence alignments and phylogenetic analysis were performed using MEGA5 (Tamura et al., 2011). For phylogeny reconstruction the neighbor joining algorithm on Poisson-corrected distances was used. Node support was estimated by the bootstrap method using 1000 replicates (Nei & Kumar, 2000). Opsin sequences identified in *A. pisum* were used to design primers for PCR amplification of aphid gene transcripts (Supplementary Table S9). The Conserved Domain Search tool (CD-Search) at NCBI was used to identify conserved domains present in aphid opsins.

[1] http://bipaa.genouest.org/is/aphidbase/acyrthosiphon_pisum/
[2] http://bipaa.genouest.org/is/aphidbase/myzus_persicae/

Table 3.1: Opsin sequences from different insect species used as queries in BLASTP searches against the *A. pisum* genome

Name[1]	Organism	Accession
Boceropsin	*Bombyx mori*	BAB63283
Rhodopsin 7	*Drosophila melanogaster*	NP_524035
UV-wavelength like opsin	*Megoura viciae*	AAG17120
Long-wavelength like opsin	*Megoura viciae*	AAG17119
Opsin 4	*Homo sapiens*	AAI43689
Rhodopsin	*Xenopus laevis*	NP_001080517
Blue-sensitive opsin	*Gryllus bimaculatus*	AEG78685
Long-wavelength opsin	*Acyrthosiphon pisum*	CAD33854

[1]Name of gene in the database

Aphids and rearing conditions

A. pisum strains YR2 (York Red 2) and GR (Gallur Rojo) were used for the different analysis. Strain YR2, collected originally in York (UK), and strain GR, collected in Gallur (Spain) are our holocyclic and anholocyclic reference strains respectively. Aphid stocks of both strains have been maintained for years in our laboratory as viviparous parthenogenetic clones on *Vicia faba* plants under long day (LD, 16L:8D) photoperiodic conditions at 18^{o}C. To study differences in mRNA expression related to photoperiodism, groups of insects from aphid strains YR2 and GR of the third larval stage (L3) were transferred to short day (SD) conditions (12L:12D). The first generation of aphids born from aphids transferred to short day conditions is referred to as SD-G1. In the case of the YR2 holocyclic strain, this SD-G1 generation consists of sexuparae which

will eventually give birth to the sexual morphs (G2 generation). For the GR strain all the generations, both under LD or SD, consist of parthenogenetic females as this strain lacks the capability to respond to photoperiod shortening.

Induction of the sexual response of YR2 at SD was confirmed in all cases by examining some individuals of the G2 generation. Sexual females were identified by detection of oviparae-specific characteristics such as thick hind tibiae (Miyazaki, 1987) and their oviparity. The presence of males, morphologically very distinct, was also recorded.

Experimental characterization of opsin transcripts by RT-PCR

For the molecular characterization and experimental validation of the opsin gene transcript predictions available at the database, a reverse transcription PCR (RT-PCR) approach was followed on aphids from the two strains described above reared under LD conditions. Aphids were synchronized allowing adult individuals to reproduce on excised *Vicia faba* leaves and selecting those that were born during a two hours window. Synchronized aphids were transferred to fresh plants and allowed to develop fully. After the final molt, on the day they started parturition, adult aphids were frozen in liquid nitrogen and kept at -80^{o}C until RNA extraction. Total RNA was extracted

using TRI Reagent® Solution following supplier's recommendations (Ambion). RNA was quantified by spectrophotometry using the NanoDrop ND-1000 (Nanodrop Technologies, Inc., Wilmington, DE, USA) and stored at -80ºC until use. Five micrograms of total RNA were used for cDNA synthesis using the Transcriptor First Strand cDNA Synthesis Kit (Roche Applied Science), primed with Oligo $(dT)_{18}$ to enrich the sample with cDNA from mature mRNAs. After PCR amplification of opsin cDNAs, amplified products were precipitated with 4M ammonium acetate and directly sequenced using the ABI Prism BigDye® terminator v3.1 Cycle Sequencing kit (Applied Biosystems) in an ABI PRISM 3700 sequencer. Chromatogram handling and processing was performed using the STADEN package (Staden et al., 1998). Aligned sequences were edited using GeneDoc and Jalview 2.8 softwares (Nicholas et al., 1997; Waterhouse et al., 2009). Secondary structure was predicted using Jpred3 (Cole et al., 2008). The mRNA sequences obtained were aligned to the genomic sequence using Unipro UGENE software (Okonechnikov et al., 2012)

Anatomical localization of opsin transcripts through PCR

As a first approach to the anatomical localization of opsin gene expression, a series of PCR reactions were performed on cDNAs obtained from dissected tissues of the aphid strain YR2 reared under LD conditions (Figures 3.2 and 3.9).

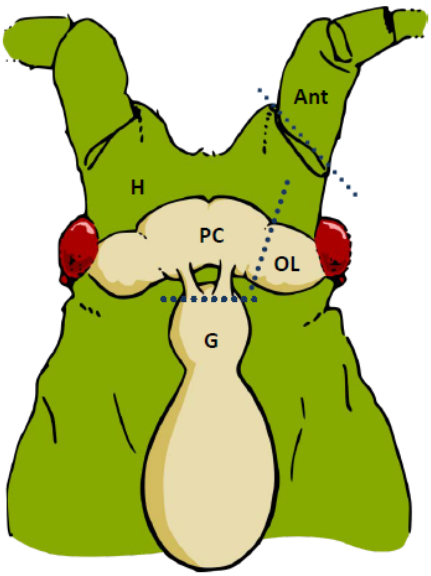


Figure 3.2: Aphid head sectioning. Schematic aphid head showing sections analysed for opsin gene expression. The head (H) sections include the carcass left after removing all the other structures (Optic Lobe plus eyes, OL; Antennae, Ant; Ganglia, G; and Protocerebrum, PC).

Around 25-30 3rd nymphal state (L3) and a similar number of adult aphids were dissected in PBS with 0.1% TritonX100. Three biological replicates of each condition were performed. Aphid heads were excised and separated samples containing antenna (Ant), optical lobe/eyes (OL), ganglia (G), protocerebrum (PC) and the remains of the head carcasses (H) were prepared (see Figure 3.2). Additionally, embryos were extracted from the abdomen of adult aphids and classified as early embryos (EE) and late embryos (LE) according to the absence or presence of visible eyes. Dissections

were performed under an Olympus SZ61 stereo microscope using Dumont nº5 and nº4 tweezers. The dissected tissues were immediately pooled in 500µl of cold TRI reagent and stored at -80ºC until RNA extraction following the manufacturer's recommendations (Ambion). Synthesis of cDNA was done using 100ng of total RNA as described above. To screen for opsin gene expression, RT-qPCR primers (see Supplementary Table S9) were used for all genes except for LWO, which was amplified using OpsCiF and ApLWOR1 primers (see Supplementary Table S9). Amplification of a fragment of the translation elongation factor 1α gene (EF1α) was used as a control of constitutive expression. PCR results were observed after electrophoresis in 2% agarose gels stained with ethidium bromide (0.5µg/ml).

Aphid central nervous system dissection

To physically localize the place of transcription of the identified aphid opsin genes, whole mount *in situ* hybridizations were performed on the central nervous system and eyes using digoxigenin (DIG)-labelled RNA probes, as seen in other studies (Velarde et al., 2005). Aphid brain extraction, probe synthesis and hybridization were performed following slightly modified protocols optimized for *D. melanogaster* (Wülbeck et al., 2007) and adapted by our group for aphids (Barberà et al., 2017). For aphid central nervous system

dissection, aphids were fixed in PFAT-DMSO (4% paraformaldehyde in 1X PBS[3], 0.1% TritonX-100, 5% DMSO v/v) over night at 4ºC to harden the tissues and preserve RNA molecules. Fixed individuals were washed 3 x 10 min in PBSTx (TritonX-100 0.1% in 1X PBS). Subsequently, dissections were carried out in glass block dishes on ice in either cold PBSTx or PBSTw (0.1% Tween 20 in 1X PBS) under a stereo microscope. The head capsules were opened with tweezers and the central nervous systems (CNS), including brain and ganglia (when possible), were extracted. The extracted CNSs were washed twice with cold 100% methanol and stored (also in methanol) at -20ºC in microcentrifuge tubes sealed with Parafilm® until the hybridization protocol was initiated.

Riboprobe synthesis

RNA probes labelled with digoxigenin (DIG) were synthesised by amplifying cDNA regions ranging from 500-900 bp for the genes of interest (see primers and amplicon sizes in Table 3.2).

PCR products were purified by ammonium precipitation and the amplicon size was checked by electrophoresis on agarose gels. Next, the purified PCR product was cloned into pGEM® -T easy vector (Promega) and used to transform DH5α *Escherichia coli* competent cells (Life Technologies). Then, colony PCR (and subsequent gel

[3]PBS: Phosphate-buffered saline. See section 3.3.2 for composition

Table 3.2: Primer combinations used to synthesize opsin riboprobes. Fragment size (in nucleotides) and position from initial methionine is provided.

Opsin Gene Name*	Primer Combination			Probe size (nt)	Position
Ap-LWO	ApLwoF2	/	OpsCiR	783	46-828
Ap-SWO1	Uv1op-FQ	/	Uv1op-R2	537	1035-1571
	Uv1op-F3	/	Uv1op-R1	517	1516-2032
Ap-SWO2	Uv2op-F1	/	Uv2op-RQ	935	65-999
Ap-SWO3	Uv3op-F3	/	Uv3op-RQ	512	144-656
Ap-C-Ops	ApPops-F2	/	ApPops-R2	720	23-742
	ApPops-F3	/	ApPops-R1	614	665-1283
Ap-SWO4	ApMops-F2	/	ApMops-RQ	973	603-1575
	ApMops-F3	/	ApMops-R2	444	1815-2258
Ap-Arth.	Artrops-F3	/	ArtrM-R6	717	405-1121

*Gene name: See text for details on gene names, section 3.3.3.

electrophoresis) was used to screen for transformants on randomly picked white colonies. Colony PCR products presenting the predicted amplicon size (see Table 3.2) were purified and sequenced from both extremes using primers SP6 (TATTTAGGTGACACTATAGAA) and T7 (GTAATACGACTCACTATAGGGC) to know the orientation of the insert. Subsequently, minipreps of selected colonies were performed to obtain high amounts of purified insert-containing plasmids using the High Pure Plasmid Isolation Kit (Roche). The templates for the synthesis of riboprobes were obtained by doing a PCR with SP6 and T7 primers on the corresponding purified plasmid and purifying the DNA fragments with the High Pure PCR Product Purification Kit (Roche). The purified PCR products were

then quantified by spectrophotometry using a NanoDrop ND-1000 (Nanodrop Technologies) and used as a template for RNA transcription. Taking into account the insert orientation, purified templates were used to synthesise DIG-labelled sense and antisense riboprobes with RNApol SP6 or RNApol T7 (Roche). Sense probes were used as negative controls. Antisense probes were used to localise gene specific mRNAs. Transcription and DIG-labelling were carried out in 0.2 ml microcentrifuge tubes following the manufacturer's protocol (Roche). DIG-labelled probes were purified with mini QuickSpin RNA columns (Roche) following the manufacturer's instructions, diluted 50% with deionised formamide and stored at -20^{o}C until use. Probe quantification was assessed by dot blot assays. Serial dilutions of the synthesized probes and a control DIG-labelled RNA of known concentration were UV crosslinked to a nylon membrane (BrightStar-Plus, Ambion). Next, the membrane was blocked with Blocking Reagent (Roche) and incubated with alkaline phosphatase (AP) anti-DIG Fab fragments (Roche). After washing the excess of antibody, NBT/BCIP substrate (Roche) was used for detection. Probe concentration was assessed by comparison of signal intensity between spots of our labelled probes with those of control RNA.

Central nervous system *in situ* hybridization

Transcript localization in the central nervous system (CNS) was carried out using the DIG-labelled RNA probes synthesized as described above. The necessary reagents and buffers where prepared as follows:

- PBS (Phosphate-buffered saline): Saline concentration was adjusted to the needs of the brain preparations to avoid tissue disruption by water invasion: NaH2PO4·2H2O 0.321 g/l; Na2HPO4·7H2O 2.07 g/l; 90g NaCl 9 g/l. pH: 7.4. For further information see Wülbeck, et al. (2007).
- PBTw: 0.1% Tween 20 in RNase free 1X PBS
- PFATw: 4% para-formaldehyde in 1X PBS, 0.1% Tween 20
- SSC (saline sodium citrate): 50% formamide and 0.1% Tween 20.
- preHB (pre-Hybridisation Buffer): 15x SSC (saline sodium citrate), 50% formamide, and 0.1% Tween 20.
- HB (Hybridisation Buffer): preHB plus 100 µg /ml heparin, 100 µg /ml salmon sperm DNA and 500 µg/ml yeast tRNA. pH 6.0.

Fixed CNSs (see above) stored in 100% methanol were gradually rehydrated in PBTw, washed in PBTw, and treated with proteinase K (3 µg/ml) for 3 min at room temperature and 1 h on ice. Proteinase digestion was blocked with 2 mg/ml glycine. Next, CNSs were refixed in PFATw for 20 min and washed 5 times in RNase free PBTw for 5 min each. Then, the samples were incubated for

5 min in a 1:1 mix of PBTw and pre-hybridisation buffer (preHB). Blocking was carried out by incubating for 2 h on a water bath at 65ºC in hybridisation buffer (HB). Next, the hybridisation was done by incubating the blocked CNSs with HB plus the corresponding probe overnight at 65ºC in a water bath. Probe concentration was determined empirically and individually, ranging from 1:100 to 1:200 dilution in the HB. Non-specifically hybridised probe was removed with the following washes at 65ºC: two washes of 15 min in preHB, three washes of 15 min each in increasing mixes of preHB:2xSSC 0.1% Tween 20 (3:2, 1:1, 2:3), a single wash of 15 min in 2xSSC 0.1% Tween 20, followed by two washes of 15 min in 0.2xSSC 0.1% Tween 20. At this point, the following washes were carried out at room temperature: three washes of 10 min each in increasing mixes of 0.2xSSC 0.1% Tween 20: PBTw transition (3:2, 1:1, 2:3), and three washes of 5 min in PBTw. Next, CNSs were blocked to avoid unspecific binding of the AP conjugated anti-DIG antibody by incubation for 2 h in PBTw supplemented with 1% Blocking Reagent (Roche) at room temperature with gentle shaking. Afterwards, the blocking buffer was replaced with fresh blocking buffer plus 1:1000 AP-conjugated anti-DIG Fab fragments and incubated overnight at 4ºC with gentle shaking.

Riboprobe detection

Alkaline phosphatase conjugated with Anti-DIG (Roche) was used to detect bound probes using either the NBT/BCIP[4] or HNPP/-Fast Red TR[5] substrates (HNPP Fluorescent Detection Set, Roche Applied Sciences). NBT/BCIP produces a blue/purple precipitate and Fast Red/HNPP produces a chromogenic red and fluorescent precipitate. Both substrates were used following the manufacturer's instructions. Times of incubation with substrates varied depending on the expression levels of the transcript being detected. The reaction was stopped by washing three times for 5 min in PBTw. Then, the tissues were immersed in increasing concentrations of glycerol (25%, 50%, 75% and 100%). Finally, the CNS preparations were mounted on microscope glass slides and sealed with nail polish.

Microscopy observation and imaging

Preparations were observed in a Eclipse E800 microscope (Nikon) and images taken with a DS-Ri1 CCD camera (Nikon) using the NIS-Elements F software v4.30.01 (Nikon).

[4]Nitro blue tetrazolium chloride/5-bromo-4-chloro-3-indolyl phosphate.

[5]4-chloro-2-methylbenzenediazonium hemi-zinc chloride salt/2-hydroxy-3-naphtoic acid-2'- phenylanilide phosphate.

Quantification of opsin gene expression through RT-qPCR

Excised adult aphid heads from YR2 and GR strains were used to compare the expression of opsin genes under different photoperiodic conditions and at different times of the day. We followed a procedure similar to that described in Cortés et al. (2008) (Figure 3.3). From each strain, aphids were sampled at 6 different time points along the day-night cycle during 1.5 days starting three hours after the lights went on (zeitgeber 3 or ZT3) and thereon taken at 6 hours intervals (i. e. ZT9, ZT15, ZT21 and on the second day at ZT3 and ZT9). Aphids reared under both long day (LD) and under short day (SD-G1) conditions were sampled (see Figure 3.3). Each sample consisted of 50 aphid heads. Three biological replicates were obtained for each condition, and for each sample three technical replicates were used. RNA extraction and cDNA synthesis were done as described above for whole aphids. RT-qPCR was performed using the StepOnePlus Real-Time PCR System, the Power SYBR Green PCR Master Mix (Applied Biosystems) and, for each opsin gene, the corresponding primers listed in Supplementary Table S9 were used.

Relative expression was expressed using the $2^{-\Delta\Delta C_T}$ method (Livak and Smitgen, 2001). From each sample, three technical replicates were analyzed. The RpL7 gene was used as an endogenous

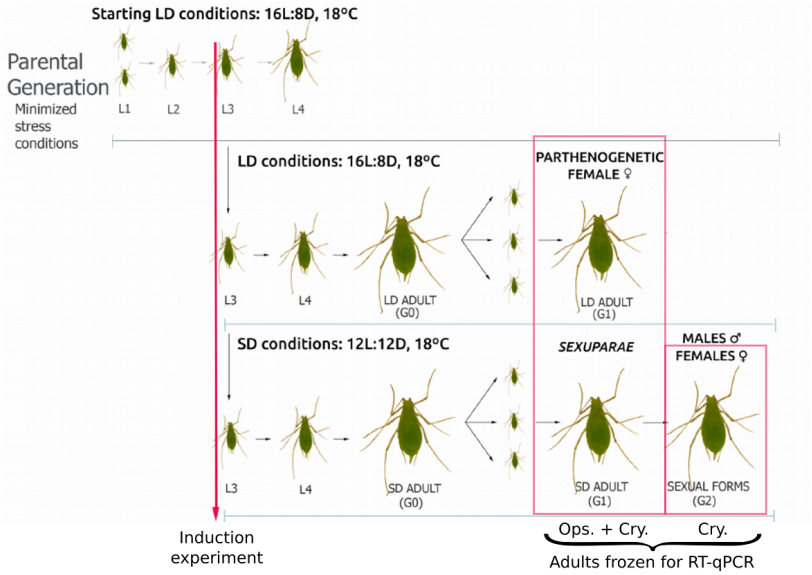


Figure 3.3: Aphid sampling for mRNA quantification. Aphids of the G0 generation where exposed to SD or LD conditions since larval stage 3. Adults of the G1 and G2 generation were frozen in liquid nitrogen and their heads were excised. For opsins (Ops.) and cryptochromes (Cry.) both parthenogenetic and sexuparae adults were used. The cryptochrome study also included adult sexual females. The anholocyclic strain used was GR, while the holocyclic strains were YR2 for opsins and LSR1 for cryptochromes (see text). L1 to L4: the four larval stages.Red boxes indicate the analyzed generations for each experiment.

control of constitutive expression (Nakabachi et al., 2005). All relative expression values were normalized to an inter-run calibrator sample (IRC). The IRC consists of an an RNA sample extracted from aphid heads and was included in every qPCR run. Two-way ANOVA was used to analyze the effects of photoperiod, strain and ZT on gene expression with the SPSS Statistics 21.0 software (IBM Corp., 2010). On those cases where ANOVA revealed a significant difference in expression between ZTs, the presence of circadian rhythmicity was tested by conducting a COSINOR analysis using the Cosinor.exe v.2.3 software (Roberto Refinetti, University of South Carolina, Salkehatchie, SC).

COSINOR analysis

The COSINOR analysis (Cornelissen, 2014) is a commonly used least-squares approach to detect a circadian rhythm and estimate its parameters. It can be used on short time series and, although it is not our case, with non-equidistant data. On those cases where ANOVA revealed a significant difference, the presence of circadian rhythmicity was tested by conducting a COSINOR analysis ($P<0,05$) with Cosinor.exe v.2.3 (Roberto Refinetti, University of South Carolina, Salkehatchie, SC).

3.3.3 Results

Identification of the opsin gene repertoire in the *A. pisum* genome

BLASTP searches on the *A. pisum* RefSeq database[6] using sequences shown in Table 3.1 as queries always rendered the same seven predicted aphid sequences as best hits (Table 3.3). To determine their orthology relationships with known opsins we carried out a phylogenetic analysis that included the seven aphid amino acid sequences and a set of visual and non-visual opsins from a wide range of organisms including insects and vertebrates representative of all opsin groups described so far (Figure 3.4). We also included opsin sequences found in *Myzus persicae*, a second aphid species whose genome was available (see Figure 3.4). All sequences analyzed cluster together with the opsins group and share similar exon distribution patterns (Figure 3.5)

Five aphid sequences grouped in a clade made up of R-opsins from other arthropods, one sequence (XP_001952294) grouped with C-type opsins (we refer to the latter as Ap-C-Ops, after *A. pisum* C-opsin) and one sequence (XP_003247319) belongs to a sister group of the R-opsins. Aphid R-type opsins clustered with sequences described as visual pigments: sequence XP_00197730 grouped with

[6]http://www.aphidbase.com/aphidbase

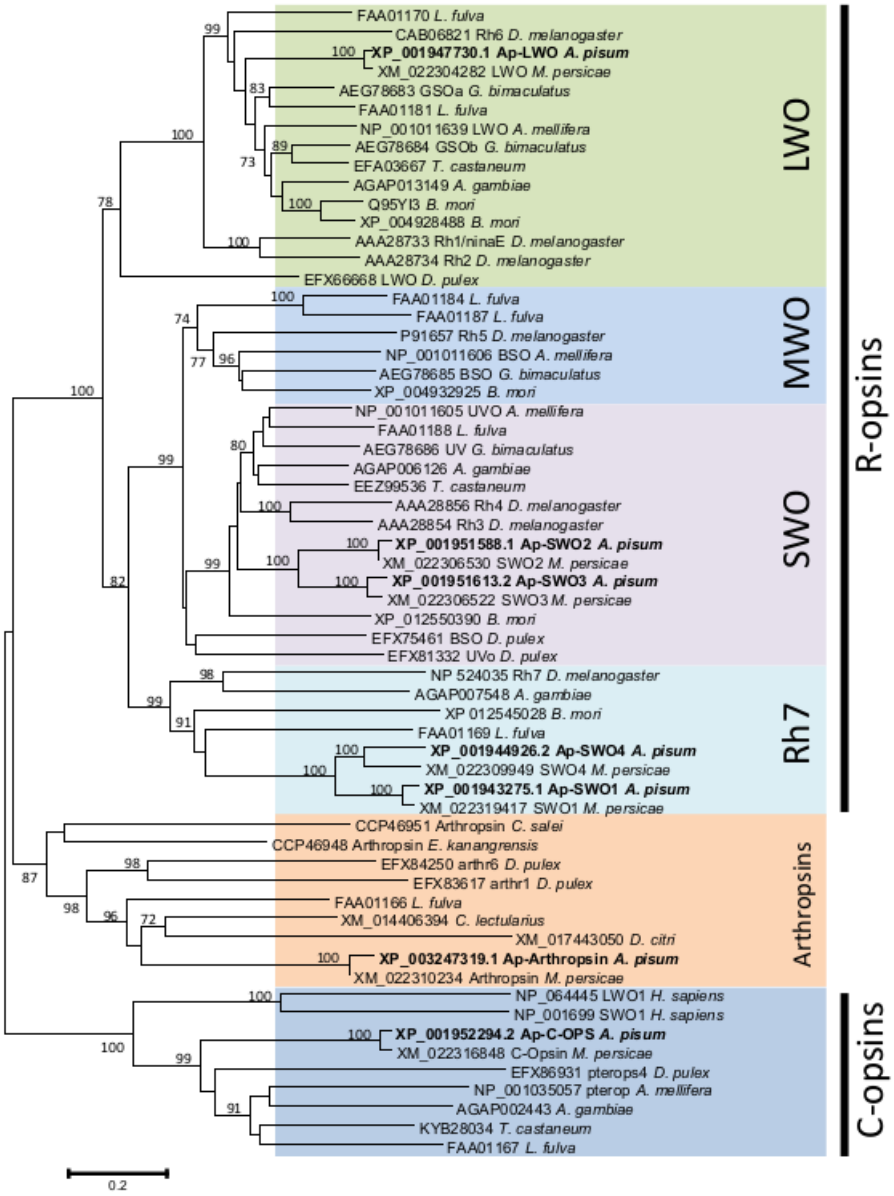

FAA01170 L. fulva
CAB06821 Rh6 D. melanogaster
XP_001947730.1 Ap-LWO A. pisum
XM_022304282 LWO M. persicae
AEG78683 GSOa G. bimaculatus
FAA01181 L. fulva
NP_001011639 LWO A. mellifera
AEG78684 GSOb G. bimaculatus
EFA03667 T. castaneum
AGAP013149 A. gambiae
Q95YI3 B. mori
XP_004928488 B. mori
AAA28733 Rh1/ninaE D. melanogaster
AAA28734 Rh2 D. melanogaster
EFX66668 LWO D. pulex
LWO
FAA01184 L. fulva
FAA01187 L. fulva
P91657 Rh5 D. melanogaster
NP_001011606 BSO A. mellifera
AEG78685 BSO G. bimaculatus
XP_004932925 B. mori
MWO
NP_001011605 UVO A. mellifera
FAA01188 L. fulva
AEG78686 UV G. bimaculatus
AGAP006126 A. gambiae
EEZ99536 T. castaneum
AAA28856 Rh4 D. melanogaster
AAA28854 Rh3 D. melanogaster
XP_001951588.1 Ap-SWO2 A. pisum
XM_022306530 SWO2 M. persicae
XP_001951613.2 Ap-SWO3 A. pisum
XM_022306522 SWO3 M. persicae
XP_012550390 B. mori
EFX75461 BSO D. pulex
EFX81332 UVo D. pulex
SWO
NP 524035 Rh7 D. melanogaster
AGAP007548 A. gambiae
XP 012545028 B. mori
FAA01169 L. fulva
XP_001944926.2 Ap-SWO4 A. pisum
XM_022309949 SWO4 M. persicae
XP_001943275.1 Ap-SWO1 A. pisum
XM_022319417 SWO1 M. persicae
Rh7
R-opsins
CCP46951 Arthropsin C. salei
CCP46948 Arthropsin E. kanangrensis
EFX84250 arthr6 D. pulex
EFX83617 arthr1 D. pulex
FAA01166 L. fulva
XM_014406394 C. lectularius
XM_017443050 D. citri
XP_003247319.1 Ap-Arthropsin A. pisum
XM_022310234 Arthropsin M. persicae
Arthropsins
NP_064445 LWO1 H. sapiens
NP_001699 SWO1 H. sapiens
XP_001952294.2 Ap-C-OPS A. pisum
XM_022316848 C-Opsin M. persicae
EFX86931 pterops4 D. pulex
NP_001035057 pterop A. mellifera
AGAP002443 A. gambiae
KYB28034 T. castaneum
FAA01167 L. fulva
C-opsins
0.2

Figure 3.4: Previous page: Neighbor joining tree built using Poisson-corrected distances on different insect and vertebrate opsin sequences including the seven pea aphid sequences identified in this report (in bold). Branch lengths measured in number of substitutions per site. Bootstrap support values higher than 70 are indicated above the branches (1000 bootstrap pseudoreplicates). Accession numbers of sequences used are indicated next to the species names. R-opsins ans C-opsins are indicated by the vertical lines. The tree was constructed using MEGA5 (Tamura et al., 2011). LWO, Long Wavelength Opsin; MWO, Medium Wavelength Opsin; SWO, Short Wavelength Opsin; Rh7, *Drosophila* Rh7-like opsins; *A. gambiae, Anopheles gambiae* (Diptera); *A. mellifera, Apis mellifera* (Hymenoptera); *A. pisum, Acyrthosiphon pisum* (Hemiptera); *B. mori, Bombyx mori* (Leidoptera); *C. lectularius, Cimex lectularius* (Hemiptera); *C. salei, Cupiennius salei* (Araneae); *D. citri, Diaphorina citri* (Hemiptera); *D. melanogaster, Drosophila melanogaster* (Diptera); *D. pulex, Daphnia pulex* (Crustacea: Cladocera); *E. kanan, Euperipatoides kanangrensis* (Onychophora: Peripatopsidae); *G. bimaculatus, Gryllus bimaculatus* (Orthoptera); *H. sapiens, Homo sapiens* (Primates); *L. fulva, Ladona fulva* (Odonata); *M. persicae, Myzus persicae* (Hemiptera); *T. castaneum, Tribolium castaneum* (Coleoptera).

high support with other insect sequences defined as long-wavelength sensitive opsins (and it was accordingly named Ap-LWO, for *A. pisum* Long Wavelength Opsin, see Table 3.3) and four sequences grouped in a clade that included short-wavelength/UV sensitive opsins from other arthropods and were given names Ap-SWO1 to Ap-SWO4 (for *A. pisum* short wavelength opsins 1 to 4, see Table 3.3).

Within the latter clade, two of the aphid sequences (sequences

Table 3.3: *Acyrthosiphon pisum* best hit sequences obtained after BLASTP searches against the Aphidbase RefSeq database using as queries opsin sequences shown in Table 3.1. Number of predicted exons, length of predicted proteins and database accessions are shown. The names we assigned to the genes were determined according to phylogenetic relations (see Figure 3.4)

Name	Exons[1]	Length[2]	ACYPI	RefSeq [3]
Ap-LWO	7	377	ACYPI009332	XP_001947730.1
Ap-SWO1	4	423	ACYPI001006	XP_001943275.1
Ap-SWO2	8	371	ACYPI004442	XP_001951588.1
Ap-SWO3	8	377	ACYPI002544	XP_001951613.2
Ap-SWO4	4	627	ACYPI005074	XP_001944926.2
Ap-Arthropsin	12	492	ACYPI32968	XP_003247319.1
Ap-C-Ops	6	386	ACYPI009397	XP_001952294.2

[1] Number of predicted exons.
[2] Length (number of amino acids) of predicted protein.
[3] RefSeq Peptide ID (NCBI).

XP_001951588 and XP_001951613 or Ap-SWO2 and Ap-SWO3 respectively) were highly similar (see Figures 3.5 and 3.6) and were located on the same genome scaffold separated by only 8,565 bp.

These two sequences grouped within a major clade formed by short-wavelength sensitive opsins from other arthropods. The other two aphid sequences included in the short-wavelength clade (sequences XP_001943275 and XP_001944926 or Ap-SWO1 and Ap-SWO4 respectively) were also highly similar but grouped with the so called rhodopsin 7 (Rh7) from *D. melanogaster* in a separate cluster (see Figure 3.4). Sequence XP_003247319 is closer to R-opsins than to C-opsins. Although it was occasionally placed in the R-group

(Eriksson et al., 2013, Hering & Mayer, 2014) it was identified as a sister group to R-opsins with a common evolutionary origin (Schumann et al., 2016). It clusters with high support to a group of highly duplicated opsin sequences identified in the genome of the crustacean *Daphnia pulex* (Colbourne et al., 2011) and also found later in the spider *Cupiennius salei*, the velvet worm *Euperipatoides kanangrensis* (Eriksson et al., 2013) and, more recently, in dragonflies (Futahashi et al., 2015). Since the *Daphnia* sequences where they were firstly identified were called "arthropsins", we designated the aphid sequence Ap-Arthropsin (Table 3.3).

A set of seven sequences highly similar to those found in *A. pisum* were also identified in the *Myzus persicae* genome (see Figure 3.4).

Experimental validation of opsin gene transcripts in *A. pisum*

Transcripts from opsin encoding genes were partially characterized in the two *A. pisum* strains analysed in the present report. The seven RefSeq predictions were used to design primers for cDNA amplification based on their predicted 5'- and 3'- UTRs (see Supplementary Table S9). We obtained cDNA sequences from all seven genes in the two *A. pisum* strains (YR2 and GR). All the sequences coincided for the most part with predictions in the database. They

were deposited in GenBank with accession numbers from MF434469 to MF434482. Analysis of the seven aphid opsin sequences revealed the presence in all of them of the 7 transmembrane (TM) domains typical of G protein–coupled receptors (GPCRs) and the conserved Lysine in the seventh transmembrane domain typically found in opsins (Figure 3.6) (Gärtner and Towner, 1995).

Other conserved motifs and residues previously reported for opsins were also found in the aphid sequences. This is the case of the counter-ion E181 in the second extracellular domain (ECDII, see Figure 3.6) (Shichida et al., 2013), the cysteines in the extracellular domains I and II (ECDI, ECDII, see Figure 3.6) (Gärtner and Towner, 1995; Hwa et al., 1999), the D/ERY motif in the intracellular domain II (ICDI) (Franke et al., 1990) and the N/DPxxY in the seventh transmembrane helix (Mirzadegan et al., 2003) (see Figure 3.6).

Predicted protein Ap-SWO4 was much longer than both the orthologous *D. melanogaster* Rh7 sequence (which is already longer than other *D. melanogaster* opsins) and the closely related aphid Ap-SWO1 sequence (see Figure 3.5). The increased length was due to the presence of an extended N-terminal region 180 amino acids longer than the Ap-SWO1 (the extra region of Ap-SWO4 is not shown in Figure 3.6 to facilitate visualisation of the alignment). This extended N-terminal region was also found in the green peach

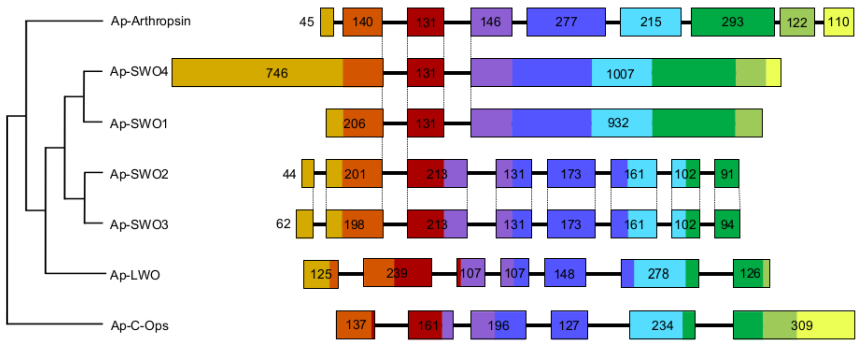


Figure 3.5: Schematic exon distribution in *A. pisum* opsin genes (coding sequence only). Numbers indicate exon length (in base pairs). Homologous regions are painted with the same color. Introns are not drawn to scale. Phylogenetic relations obtained from Figure 3.4

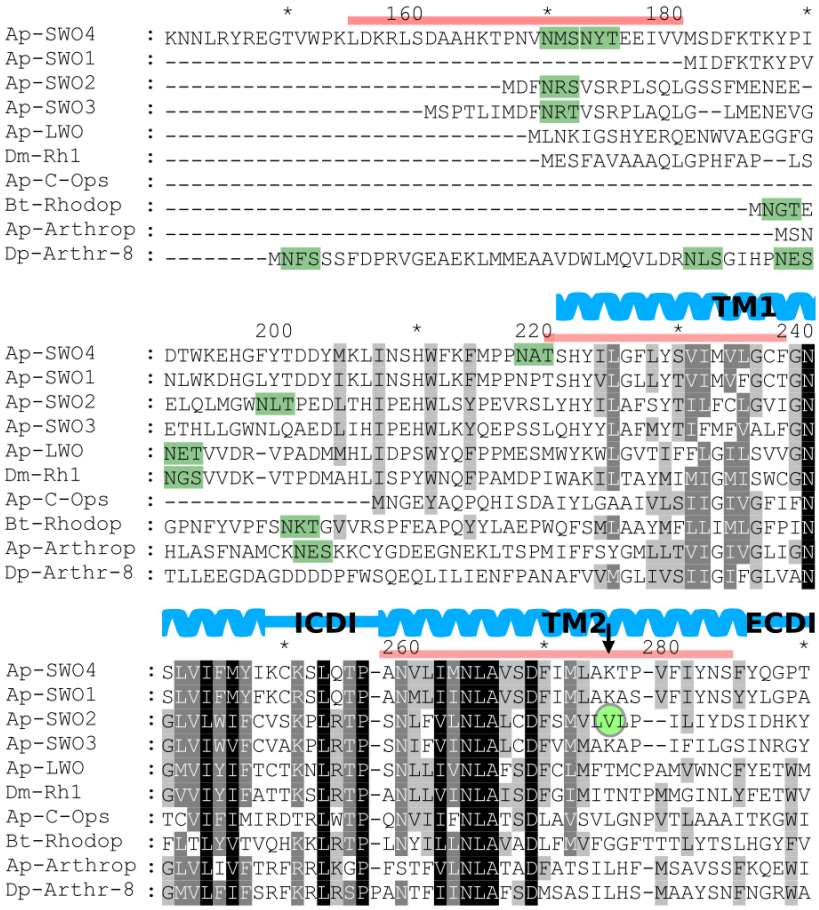
aphid *M. persicae*, according to BLAST searches, and could form extra alpha helixes in the protein. Ap-SWO1 and Ap-C-Ops both lack N-glycosylation sites widely conserved in opsins (Murray et al., 2009). Also, a QAKKMNV motif, which is highly conserved in the third intracellular loop in invertebrate opsins (Gao et al., 2000; Townson et al., 1998), and which has been associated with G-protein binding (Gao et al., 2000), is absent in 4 of the aphid proteins: Ap-SWO1, Ap-SWO4, Ap-C-Ops and Ap-Arthropsin. Consistent with previous data (Gärtner and Towner, 1995), the aphid opsin putatively sensitive to visible light (Ap-LWO) has a tyrosine in the third transmembrane domain, three residues apart from a conserved cysteine, while putative aphid UV-sensitive opsins (Ap-SWO2 and Ap-SWO3) possess a phenylalanine instead. Of the rest of aphid

Table 3.4: Non-synonymous sequence polymorphisms found in YR2 and GR strains opsin transcripts.

Gene	nt position from ATG	mRNA			Protein**		
		YR2	GR	LSR1*	YR2	GR	LSR1*
Ap-SWO4	212	AGT	AGT/AAT	AGT	S	S/N	S
	250	ATT	ATT/GTT	ATT	I	I/V	I
	388	CCC	CCC/TCC	CCC	P	P/S	P
Ap-C-Ops	971	CAG	CAG	CCG	Q	Q	P

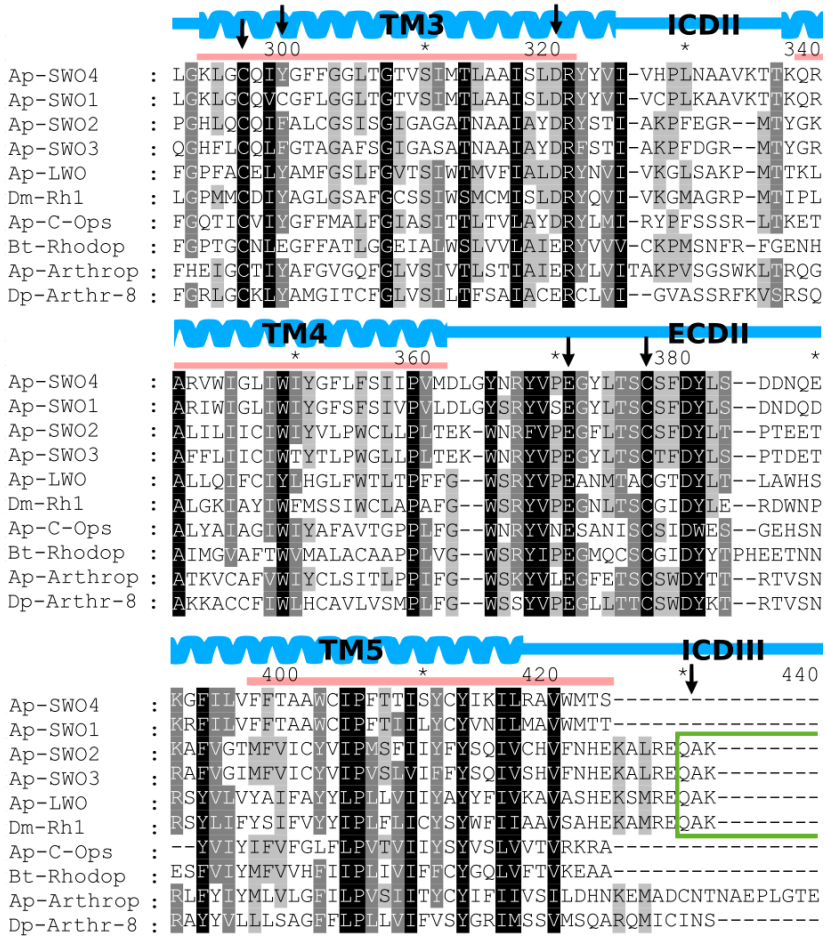
*LSR1 strain sequence of the predicted gene in the Aphid Genome Project database.
**Corresponding aminoacid in the protein sequence.

opsins, Ap-SWO4, Ap-C-Ops and Ap-Arthropsin have a tyrosine, while the other Rh7-like opsin, Ap-SWO1, has a cysteine at that position (see Figure 3.6). Among the four aphid opsins clustering with other arthropod UV/short wavelength sensitive proteins, Ap-SWO2 replaced a lysine conserved in the position corresponding to the bovine rhodopsin G90 in TMII, typical of UV sensitive opsins (Salcedo et al., 2003), for a valine. When sequences from the two studied aphid strains were compared, 39 polymorphic sites were detected (some of them heterozygous). 35 of the identified nucleotide polymorphisms corresponded to silent changes. All non-synonymous changes (four) were found in heterozygosis: three positions in Ap-SWO4 were localized in the extended N-terminal region described above, and a single non-synonymous nucleotide polymorphism was identified in Ap-C-Ops (see Figure 3.6 and Table 3.4).



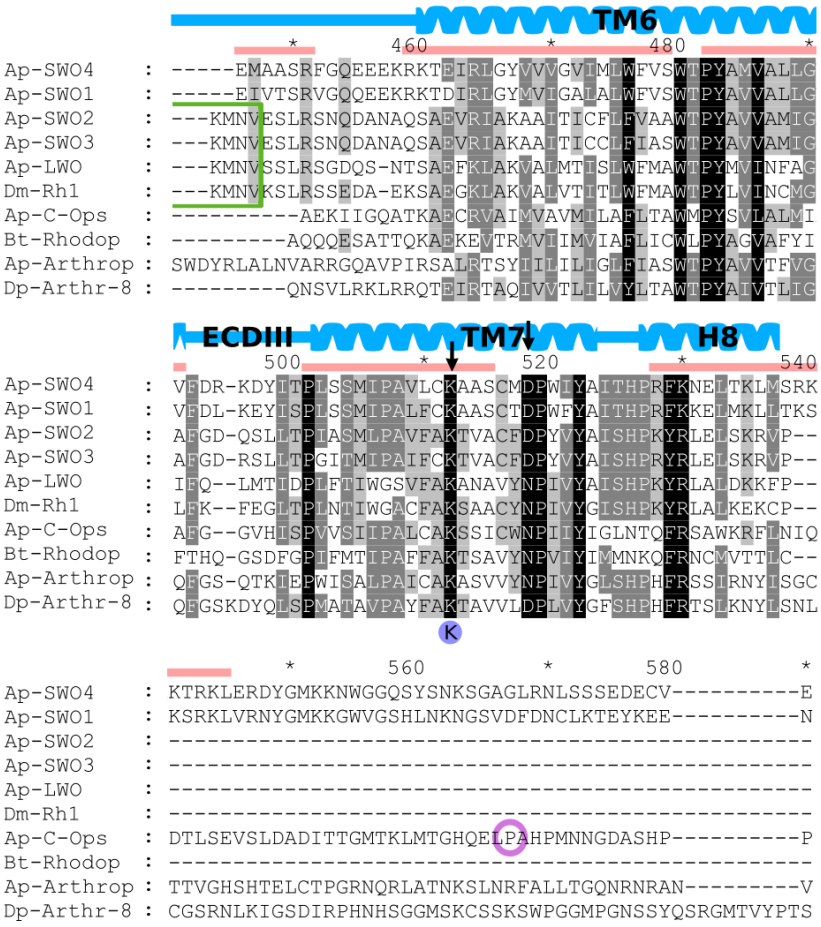
(a) Continues in next page

Figure 3.6



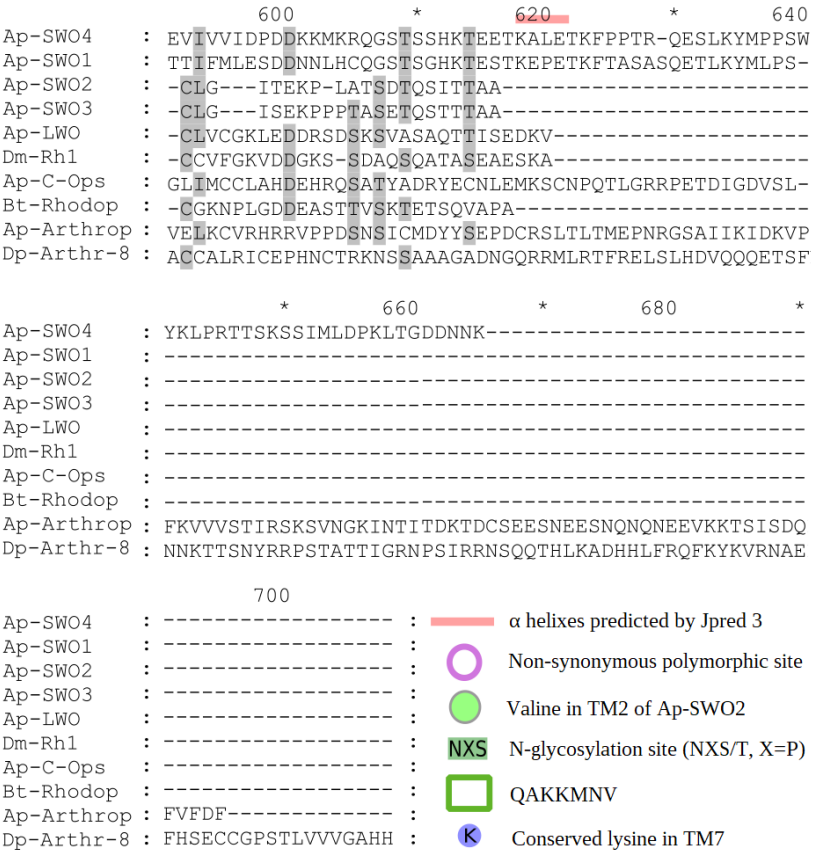
(b) Continued from previous page

Figure 3.6



(c) Continued from previous page

Figure 3.6



(d) Continued from previous page

Figure 3.6: **(a)** to **(d)**: Alignment of *Acyrthosiphon pisum* opsin protein sequences. Opsins from other species were included for reference: *Drosophila melanogaster* Rh1 (Dm-Rh1), *Bos taurus rhodopsin* (Bt-rhodop) and *Daphnia pulex* Arthropsin 8 (Dp-Arthrop-8). Residue numbering corresponds with the Ap-SWO4 sequence (trimmed to facilitate visualisation). Secondary structure indicated above the protein sequence in blue (from bovine rhodopsin) and red (predicted by Jpred3). TM1-7, Transmembrane domains 1 to 7; ICD, Intercellular Domains; ECD, Extracellular Domains; H8, Helix 8. Arrows point at residues referred to in the text. Relevant features are detailed in the legend. QAKKMNV is a widely conserved motif.

Quantification of opsin gene expression

When global expression under the two studied photoperiods was compared with ANOVA tests, each of the seven opsin genes revealed significantly higher expression under short-day (SD) conditions than under long-days (LD) in aphids of the reference holocyclic strain YR2. The values range from 1.30 to 3.69 times higher expression than LD in Ap-SWO1 and Ap-SWO3 respectively (Figure 3.7 and Supplementary Figure S2). Contrarily, in the GR strain, only two opsin genes showed significantly higher expression under SD conditions. The recently duplicated Ap-SWO1 and Ap-SWO4 opsins showed 1.62 and a 2.75 times higher expression than LD respectively when compared to SD (see Figure 3.7). The other five opsin genes showed no differences in levels of expression between LD and SD reared aphids in the GR strain. The latter five opsins showed similar levels of expression in GR under both photoperiods to those observed under LD conditions for the YR2 strain (see Figure 3.7). Ap-SWO1 expression levels were particularly low, especially in the GR strain (notice the scale size in Figure 3.7). When expression levels of opsin genes were compared along the day-night cycle by sampling aphids at six different time points (ZTs) during 1.5 days (see Materials and Methods), ANOVA analysis revealed significant differences in expression between ZTs in some opsin genes only in the

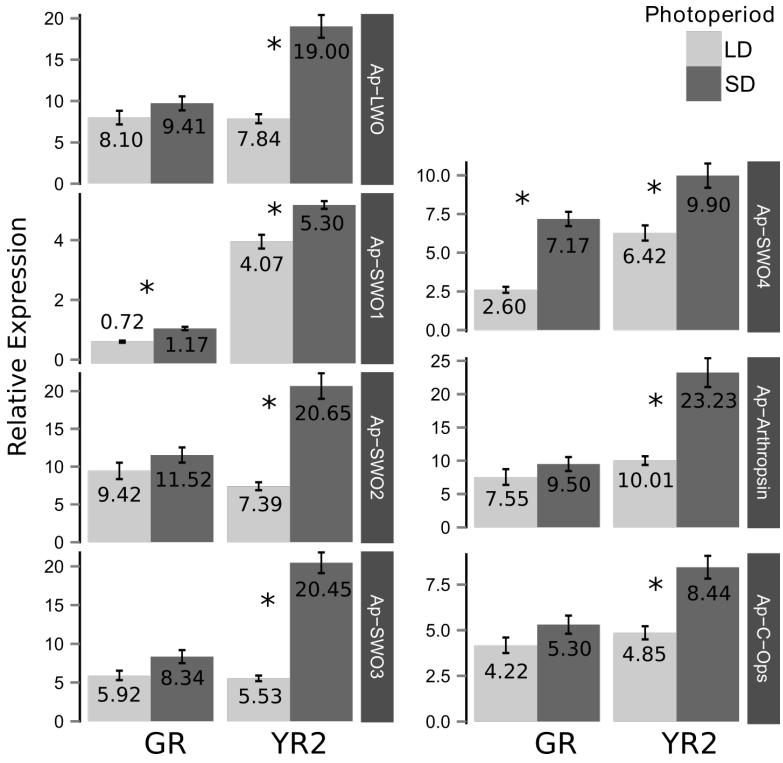


Figure 3.7: Relative expression of opsin genes in heads of adult aphids from two different strains reared under different photoperiods. Mean values are shown. YR2 and GR correspond to the holocyclic and the anholocyclic strains respectively. In the case of the YR2 strain SD aphids correspond to the G1 generation (sexuparae). Error bars represent the standard error of the mean (SEM). Significant differences in gene expression after ANOVA test are indicated by * $p<0.01$.

YR2 strain (i.e. Ap-C-Ops under LD conditions and Ap-Arthropsin and Ap-SWO4 under both photoperiodic conditions) (see Figure 3.8 and Supplementary Figure S2). However, the COSINOR analysis revealed no significant rhythmicity in the expression of these genes (see Materials and Methods).

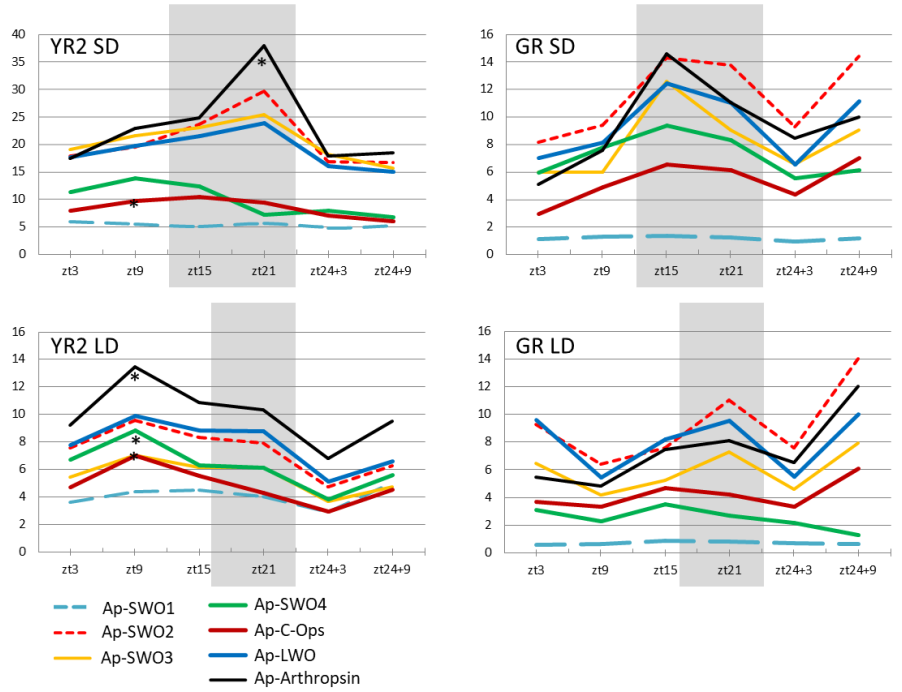


Figure 3.8: Relative expression of opsin genes in two *A. pisum* strains under two different photoperiods along 1,5 day/night cycles. Correspondence between line color and gene is indicated below the graphs. YR2: York Red 2 strain; GR: Gallur Rojo strain; ZT: zeitgeber time; SD: Short day; LD: Long day; Ap-SWO1 to 4: *Acyrthosiphon pisum* Short Wavelength Sensitive Opsins 1 to 4; Ap-C-Ops: C Type Opsin; Ap-LWO: Long Wavelength Sensitive Opsin (see the text for details on gene names). Periods of darkness are shaded. Asterisks mark those timepoints that showed significant differences in ANOVA tests ($p < 0.05$) in some of the opsin genes studied. These differences are detailed in Supplementary Figure S2. Error bars have been omitted for clarity.

Localization of opsin gene expression through PCR amplification

PCR amplification of opsin transcripts from dissected aphid sections (see Materials and Methods) suggested localized (and probably highly specific) expression of some opsin genes (see Figure 3.9). In particular, Ap-SWO2 and Ap-SWO3 were amplified mainly in Eyes/Optic lobe preparations (OL) both in adult aphids and in L3 nymphs. No amplification of these two genes was obtained from RNA preparations from the other tissues analyzed. Ap-LWO showed a similar distribution but it was also amplified in L3 head carcasses and in late embryos (LE). The other four opsins (Ap-SWO1, Ap-SWO4, Ap-Arthropsin and Ap-C-Ops) could be amplified from all sections obtained from aphid brains (OL, PC and G). Ap-SWO4 and Ap-Arthropsin were also amplified from embryo sections (EE and LE, see Figure 3.9). Ap-SWO1 was amplified from almost every tissue analyzed, being the only opsin transcript detected in antenna. PCR amplification from early embryos showed that Ap-SWO1, Ap-SWO4 and Ap-Arthropsin expression begins in the early phases of embryonic development.

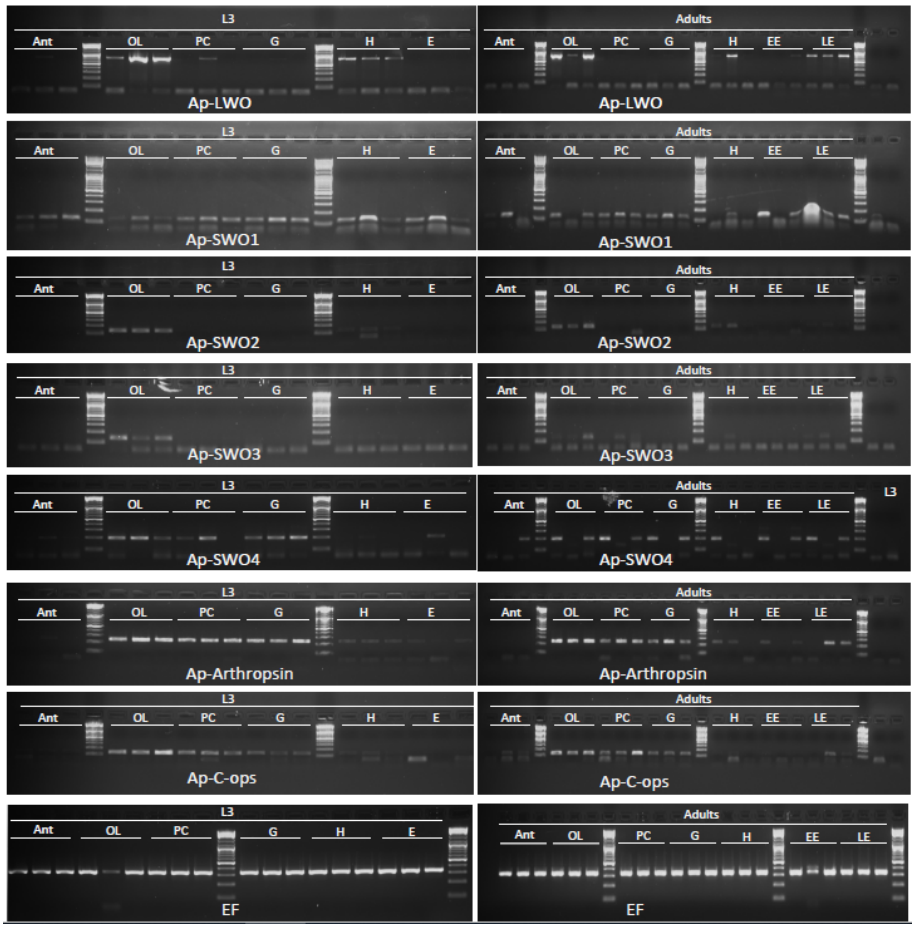
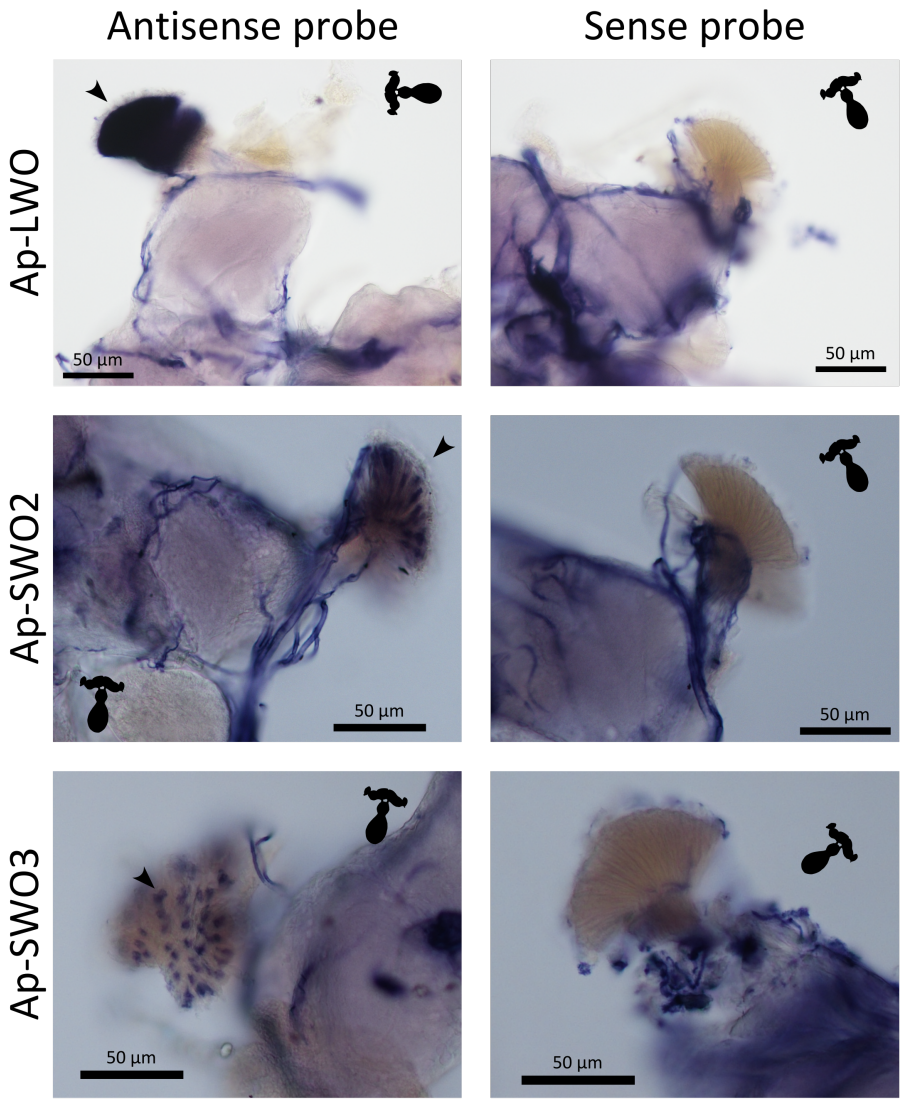


Figure 3.9: PCR amplification of opsin gene fragments from adult head sections shown in Figure 3.2 and from L3 nymphs. For each gene and tissue, three biological replicates were done. Ant, Antennae; OL, Optic lobe plus eyes; PC, Protocerebrum; G, Ganglia; EF, Translation Elongation Factor 1α; H, Head carcass. For Materials and Methods for details in the classification of E, EE and LE (Embryo, Early embryo and Late Embryo).

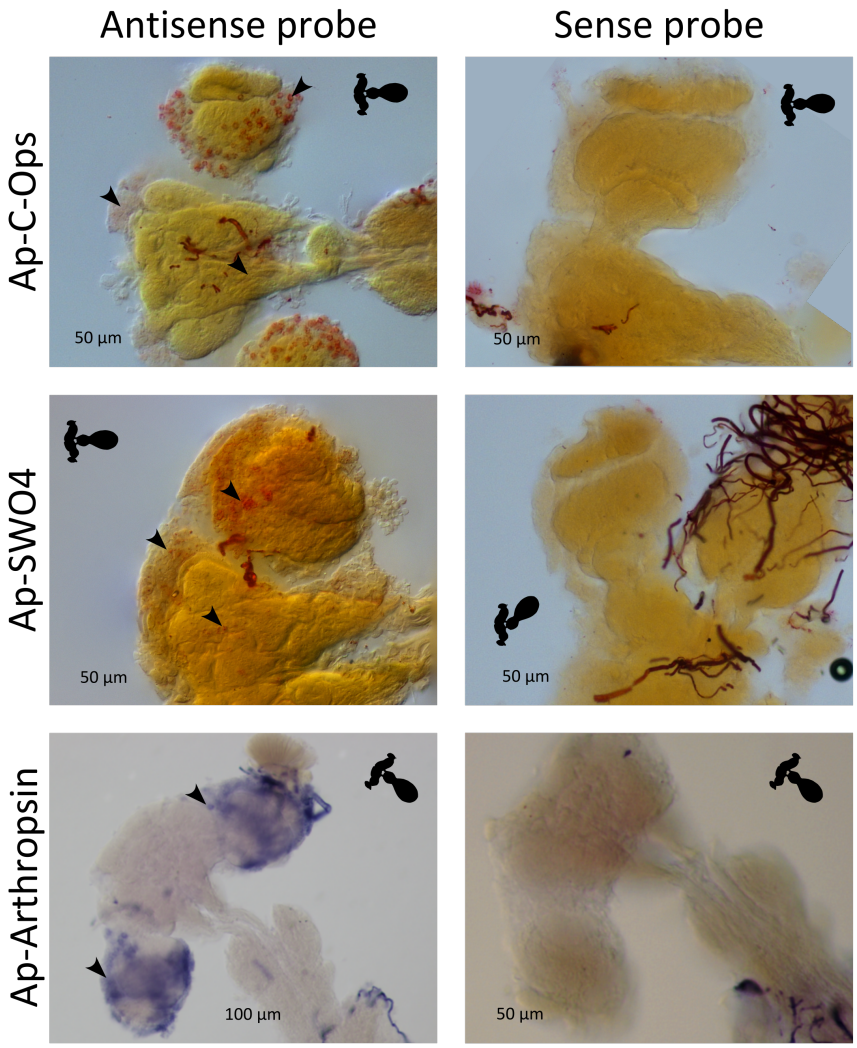
Localization of opsin gene expression by *in situ* hybridizations

Whole mount *in situ* hybridizations on aphid central nervous system and eyes revealed the presence of Ap-SWO2, Ap-SWO3 and Ap-LWO transcripts exclusively in the aphid compound eyes (Figure 3.10). Ap-LWO showed a strong widespread expression in the eyes, covering every photoreceptor cell, whereas Ap-SWO2 and Ap-SWO3 transcripts were dispersed in separate cells or cell groups. Ap-Arthropsin was only detected in the optic lobes (in the cortex of the lamina, medulla and lobula) and in individual cells in the interception between the protocerebrum and the optic lobes (see Figures 3.10 and Supplementary Figure S3). Ap-C-Ops was clearly visible in several individual cells in the optic lobe (mainly in the medulla and lobula) and presented a mild signal in the dorsal region of the protocerebrum (Figures 3.10 and Supplementary Figures S4 and S5). Ap-SWO4 was detected in the optic lobes and in specific cells of the ventral protocerebral region. We were unable to detect Ap-SWO1.



(a) Continues in next page

Figure 3.10



(b) Continued from previous page

Figure 3.10: *In situ* localisation of opsin gene transcripts in adult *A.pisum* central nervous system and eyes. **(a)**: Results for Ap-LWO, Ap-SWO2 and Ap-SWO3. **(b)**: Results for Ap-C-Ops, Ap-SWO4 and Ap-Arthropsin. Ap-SWO4 and Ap-C-Ops detected with Fast Red/HNPP substrate, the rest with NBT/BCIP (See Materials and Methods). Arrows point to specifically stained regions where antisense probes were bound to their matching mRNA. A representation of the central nervous system is included in each image indicating the orientation of the preparation

3.4 Cryptochromes

3.4.1 Introduction to cryptochromes

As mentioned in previous sections, apart from opsins, animals have another photosensitive protein family that could be candidate to be involved in the photoperiodic response: the cryptochromes (Terakita, 2005; Michael et al., 2017). Cryptochromes are known to play a crucial role in other insects' circadian clocks (Emery et al., 1998; Yuan et al., 2007; Michael et al., 2017). They are blue light-sensitive proteins non-covalently bound to a FAD molecule[7]. They are related to the photolyase family, a group of proteins in charge of repairing light-induced DNA damage. Despite their similarities with photolyases, cryptochromes have lost the ability to repair DNA but have evolved to perform other functions such as photoreception, magnetoreception or as transcription factors (reviewed in Michael et al., 2017). At least two types of cryptochrome encoding genes can be found in insects: the *Drosophila*-like cryptochrome (*dCry* or *cry1*) and the mammalian-type cryptochrome (*mCry* or *cry2* in insects[8]). While *D. melanogaster* lacks the *cry2* gene, bees and

[7]Flavin Adenine Dinucleotide

[8]Mammals have two copies of their *cryptochrome* gene. These copies are referred to as *Cry1* and *Cry2* (with uppercase), which generates confusion when talking about cryptochromes. In this work we refer to *Drosophila* or mammalian-like *cryptochrome* genes as *cry1* and *cry2* respectively (with lower case)

beetles lack *cry1*. However, butterflies and aphids have both types of cryptochrome (*cry1* and *cry2*) (Yuan et al., 2007; Cortés et al., 2010) (see Figure 1.10 in the General Introduction). *Drosophila*'s CRY1 works as a photoreceptor (Emery et al., 2000). After contact with light at dawn, it binds to TIMELESS (TIM, a central circadian clock protein, see section 1.3.1) and targets it for the ubiquitinase protein JETLAG (JET) marking it for degradation (Figure 3.11) (Koh, 2006).

In *Drosophila*, TIM degradation is fundamental for clock resetting every morning. This process synchronizes the clock with the outer light conditions, being part of a series of feedback loops together with the product of gene *period* (*per*) (Hardin, 2005). While CRY1 protein and mRNA levels fluctuate daily in *D. melanogaster* (Emery et al., 1998), transcript levels remain constant in some butterflies, rotifers, annelids and mosquitoes (Zhu et al., 2008; Gentile et al., 2009; Rund et al., 2011; Zantke et al., 2013; Meuti et al., 2015; Rund et al., 2016). The mammalian-type cryptochrome CRY2 seems to have lost photosensitivity, but acts as a relevant transcription factor, interacting actively with the central oscillator elements (Yuan et al, 2007; Mei et al., 2015, Helfrich-Forster,2019). In butterflies it inhibits transcription mediated by the products of genes *clock* and *Cycle* (*clk* and *Cyc*, respectively) coordinated, most probably, with PER (Zhu et al., 2005)(see Figure 1.10). The aphid clock is

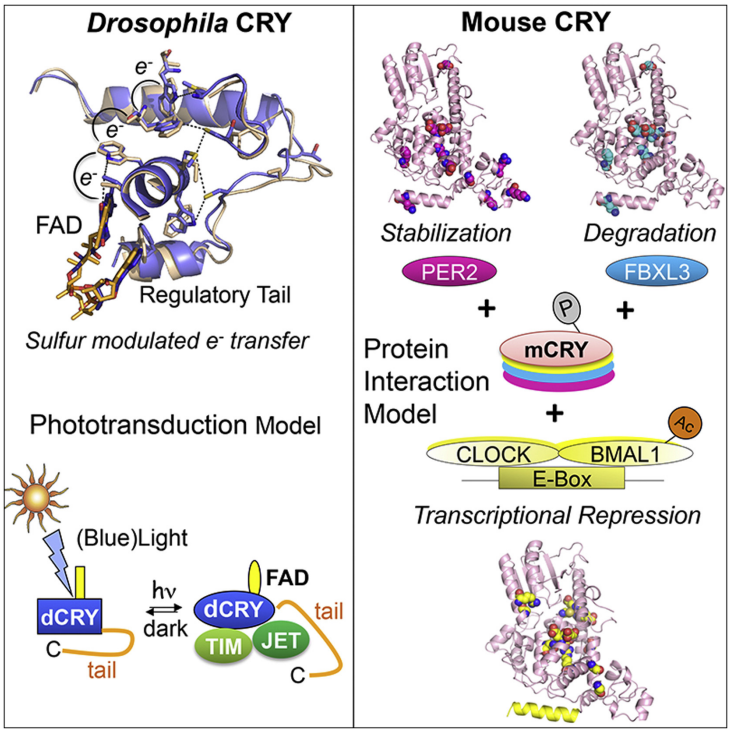


Figure 3.11: *Drosophila* and mammal-type cryptochromes, represented by the orthologs in fly and mouse. Aphids have both types. *Drosophila*-like (dCRY or CRY1) is involved in phototransduction: when exposed to light, conformational changes in the tail enable interaction with TIM and JET proteins to target TIM for degradation. In mammals, the cryptochrome (mCRY or CRY2) is a light-independent transcriptional repressor that inhibits CLOCK/BMAL dependant transcription. BMAL is the mammalian homolog of CYCLE. mCRY repressor activity depends on its interaction with PERIOD (PER2), FBXL3 (involved in mCRY polyubiquitination), BMAL acetylation (which facilitates recruitment of mCRY) and mCRY phosphorilation. Taken from Czarna et al. (2013)

of particular interest. As mentioned in the General Introduction, there are, at least, three types of clocks in insects based on the combination of *cryptochrome* genes in their genome (see Figure 1.10). As already mentioned, some insects only have one *cryptochrome*, either *cry1* or *cry2*, while others have both. Like butterflies, *A. pisum* has both *cry1* (Aphidbase accession ACYPI005768) and *cry2* genes homologs. Gene duplications are frequent in aphids and this is also true for *cry2* which is present in two copies, *cry2-1* and *cry2-2* in the aphid genome (Aphidbase accessions ACYPI006584 and ACYPI087167) (Cortés et al., 2010). As mentioned above, CRY1 is known to interact with TIM and JET to degrade TIM in *Drosophila*. Aphids, however, lack a JET homolog (Cortés et al., 2010). Moreover, similar to transcription factors at the core of the aphid clock (PER, TIM, CYC and CLK), both copies of CRY2 in *A. pisum* are highly divergent and are evolving faster than the corresponding homolog proteins in other insects.

As already discussed, it has been proposed that molecular elements of the circadian clock can also play a role in the regulation of photoperiodic responses in insects (Vaz Nunes & Saunders, 1999). Although some of those molecular elements like PER and TIM have already been proven to be related to the seasonal changes (Majercak et al., 1999; Shafer et al., 2004; Cortés et al., 2010) and to be expressed in aphid brain regions relevant for the photoperiodic re-

sponse (Barberà et al., 2017), the nature of the main photoperiodic photoreceptor (or photoroceptors) remains unknown in aphids and insects in general. In this regard CRY1, as a photosensor (at least in *Drosophila* and other insects) is, along with other photoreceptors, such as opsins (Collantes-Alegre et al., 2018 and see section 3.3.1) one of the candidates to perform this function in aphids. On the other hand, it has been shown that *cry2* homologs in other organisms don't seem to have photosensitive properties (Yuan et al, 2007; Mei et al., 2015, Helfrich-Förster, 2019). Rather, in other insects CRY2 protein has regulatory functions as a transcription factor as a co-repressor with PER. Photosensitivity of CRY2, however, might not be discarded yet, as Yuan and colleagues (2007) themselves propose, it is possible that photosensitivity can not be detected in cell cultures but still be a property of CRY2. It has been shown that CRY2 has a role in the the light-dependent biological compass of the cockroaches (Bazalova et al., 2016). Caution is also advised with this result as it is possible that cryptochromes are necessary but they might not be the photoreceptor molecule in light-dependant magnetoreception (Hore et al., 2016).

In the present thesis, we have mainly concentrated on the *cry1* aphid homolog, studying transcript levels along the day/night cycle and under different photoperiodic conditions. In addition, differently from previous studies, in which transcript quantification was limited

to parthenogenetic aphid females only (Barberà et al., 2013; Barberà et al., 2017), in this study we also included sexual females (oviparae). Sexual females are of particular interest in circadian studies, as they show clearer circadian behaviours or characteristics than parthenogenetic females. Parthenogenetic females have a very sedentary life and circadian behavior proves difficult to study. Oviparae of some aphid species, on the other hand, exhibit circadian patterns in pheromone emission (Eisenbach et al., 1980; Thieme et al., 1996; Jeon et al., 2003), one of the few rhythmic behaviors reported in aphids together with movement towards host plant (Narayandas et al., 2006). In addition, we have attempted to localize the site of expression of the two types of cryptochromes in the aphid brain, *cry1* and *cry2*. We aimed to get some evidence on their involvement in the photoperiodic response and/or the circadian clock by determining their expression pattern in the brain. Moreover, we try to disentangle the confusing terminology and equivalences found in the database for the homologs of *cry 2* present in the aphid genome.

3.4.2 Materials and methods

The study of cryptochromes in the strains of interest was carried out following methods described in section 3.3.2 for opsins. There are, however, some relevant differences worth being addressed:

- The YR2 strain was not available for experimental purposes when the project was being set up, thus another holocyclic strain was used: LSR1, collected originally near Ithaca, New York. The same anholocyclic strain as in the opsin study was used: GR, collected in Gallur (Spain).
- In this experiment the expression was also quantified by RT-qPCR in females of the sexual generation. See Figure 3.3
- More time points (ZTs) were sampled, thus we analyze not 6 but 8 ZTs covering 2 complete days

The use of longer periods of time (two days) allows us to obtain more detailed information for the study of an eventual circadian rhythm. Moreover, the inclusion of oviparae provides us with a different morph of the same clone which might present variability in gene expression due to its differences in behaviour and physiology.

Aphids

Adult aphids were used for the different analysis. Stocks of both aphid strains have been kept for years in our laboratory as viviparous parthenogenetic clones on *Vicia faba* plants under long day (LD, 16L:8D) conditions at 18^{o}C. To study differences in mRNA expres-

sion related to photoperiodism, groups of insects from aphid strains LSR1 and GR of the third larval stage (L3) were transferred to short day (SD) conditions (12L:12D) (see Figure 3.3). The first generation of aphids born from aphids transferred to short day conditions is referred to as SD-G1. In the case of the LSR1 holocyclic strain, this SD-G1 generation consists of sexuparae which gave birth to the sexual morphs (G2 generation). For the GR strain all the generations, both under LD or SD, are made of parthenogenetic females as this strain lacks the capability to respond to photoperiod shortening. The induction of the sexual morph in the LSR1 strain was confirmed by observation of the G2 generation.

Quantification of cryptochrome gene expression through RT-qPCR

Given that the photoperiodic signal in aphids is sensed through the head (see section 1), excised adult aphid heads from LSR1 and GR strains after their final moult were used to compare the expression of the *cry1* gene under different photoperiodic conditions and at different times of the day. We followed a procedure similar to that used by Cortés et al. (2008) and described before for opsin characterization in section 3.3.2. From each strain, aphids were sampled at 8 different time points along the day-night cycle during 2 days starting three hours after the lights went on (*zeitgeber* 3 or

ZT3) and thereon taken at 6 hours intervals (i. e. ZT9, ZT15, ZT21 and the same points for the second day). Synchronized adult aphids reared under both long day (LD-G1) and under short day (SD-G1) conditions (Cortés et al., 2010) were sampled, starting the following day after the final moult. Three biological replicates were obtained from each ZT. Each sample consisted of 20 aphid heads. The sexual G2 generation, was kept under SD conditions and sexual females were sampled identically after the final moult. The heads were excised in glass embryo dishes under a stereo microscpe and total RNA was extracted immediately using TRI Reagent® Solution (Ambion) and the Direct-Zol RNA extraction Kit (ZYMO) following supplier's recommendations. RNA was quantified by spectrophotometry using the NanoDrop ND-1000 (Nanodrop Technologies, Inc., Wilmington, DE, USA) and stored at -80ºC until use. Total RNA was used for cDNA synthesis using the Transcriptor First Strand cDNA Synthesis Kit (Roche Applied Science), primed with Oligo $(dT)_{18}$ to enrich the sample with cDNA from mature mRNAs. Primer design for PCR and riboprobe synthesis was based on the *cry1* sequence validated by Cortés et al., (2010). RT-qPCR was performed using the AriaMx Real-Time PCR System (Agilent), and the SYBR qPCR Master Mix (UVAT-Bio). See primers used for RT-qPCR in table 3.5. Similar to opsin analysis (see section 3.3.2), relative expression for each sample was calculated using the ΔΔCt (threshold cycle) method (Livak

Table 3.5: Primers used in *cryptochrome* genes quantification, characterization and riboprobe synthesis

Primer	Sequence (5' → 3')
apCRY1-F1	AGGCCCCTTATATGTAGTGAC
apCRY1-F2	ACGGTAATTATGACTGTCGCCG
apCRY1-F3§	GGCCTAAGGTTGCACGACAACCCAG
apCRY1-F4†	ATGCTTATCCATAAGAAAG
apCRY1-AltF1	TATTGTGTTAACAGTGCGGATGT
apCRY1-AltF2	AATGTCCACCACCGTCAATGCAC
apCRY1-R1§	CGCATCGGATTGTATCCATACCA
apCRY1-R2†	GTACCCTGTTCTCGTTCGTCAT
apCRY2-F3†	AACACACCGTCCACTGGTTCCG
apCRY2-R2†	ATTGCAACCTTATCAATGCTTCC
apCRY2-F5†	CAACCACCACTTTCGCTGCACGG
apCRY2-R8†	TTCAATGCTGGAATGTATCTTC

All primers were used for transcript characterization. §, primers used for quantification by RT-qPCR; †, primers used for probe synthesis.

and Schmittgen, 2001)

Transcript characterization of *cry1*

The cDNA obtained and analyzed by quantitative PCR was also used to characterize the transcripts. Different primer combinations were used (see table 3.5). Primers were designed based on the available mRNA sequences in the database: NM_001171061.1 (RefSeq mRNA accession) and FN377569.1 (NCBI accession for cDNA sequenced in our group by Cortés et al., 2010) and the am-

plicons were observed in agarose gel electrophoresis and sequenced. Amplified products were precipitated with 4M ammonium acetate and directly sequenced using the ABI Prism BigDye® terminatior v3.1 Cycle Sequencing kit (Applied Biosystems) in an ABI PRISM 3700 sequencer. Chromatogram handling and processing was performed using the STADEN package (Staden et al., 1998), mRNA was aligned to genomic sequences using Unipro UGENE software (Okonechnikov et al., 2012)

Localization of cryptochrome expression by *In situ* hybridization

To physically localize the place of transcription of the aphid cryptochrome genes *cry1* and *cry2*, whole mount *in situ* hybridizations were performed on dissected aphid central nervous systems using digoxigenin (DIG)-labelled RNA probes (see section 3.3.2). One probe was synthesized to detect *cry1*, while *cry2* proved harder to detect and two probes were used simultaneously to improve the detection signal. As in the opsin characterization experiment, aphid brain dissection, probe synthesis, hybridization and visualisation were performed following protocols described by Barberà et al. (2017). The riboprobe for *cry1* mRNA was synthesized from an amplicon 746 bp long (see primers in Table 3.5). The riboprobes for *cry2* were not specific to *cry2-1* nor to *cry2-2*, as they bind to com-

mon regions with high similarity (∼98%). ApCry2F3 - ApCry2R2 primer combination generated a probe 663 bp long, and 649 out of 663 positions were identical between *cry2-1* and *cry2-2* (97.89%). ApCry2F5 - ApCry2R8 primers generated a probe 470 bp long (462 out of 470 identical positions between gene copies *cry2-1* and *cry2-2*, 98.30%). To avoid cross-hybridization, probes for *cry1* and *cry2* genes were designed from non-conserved regions between these genes. Similarly to the procedure performed for opsins (see section 3.3.2), for the detection of hybridised probes we used alkaline phosphatase conjugated with Anti-DIG (Roche) and HNPP/Fast Red TR substrates (HNPP Fluorescent Detection Set, Roche Applied Sciences).

3.4.3 Results

A phylogenetic analysis of the *Acyrthosiphon pisum* cryptochromes found in the Aphidbase (accessions numbers ACYPI005768 for *cry1* and ACYPI006584 and ACYPI087167 for *cry2*) shows the position of these genes in the cryptochrome/photolyase superfamily (Figure 3.12). The photosensitive *cry1* groups together with *Drosophila*'s *cry1* while both copies of *cry2* are located in the mammalian-type group, together with human sequences.

Transcript characterization of *cry1*

Full transcript characterization was only successful for *cry1*, while *cry2* was only partially characterized due to the presence of duplications (*cry2.1* and *cry2.1*) and complex alternative splicing.

PCR amplification of *cry1* mRNA was performed using several primer combinations (see Table 3.5) designed using sequences available in the Aphidbase and NCBI databases. Based on mRNA sequences NM_001171061.1 and FN377569.1, which differ in the exon distribution in the 5' region (see Figure 3.13), the *cry1* gene model has 12 exons (Figure 3.13), with a start codon in exon 4 and a stop codon in exon 11. We have amplified, sequenced and aligned *cry1* cDNA with the genomic sequence of *A. pisum*, confirming the

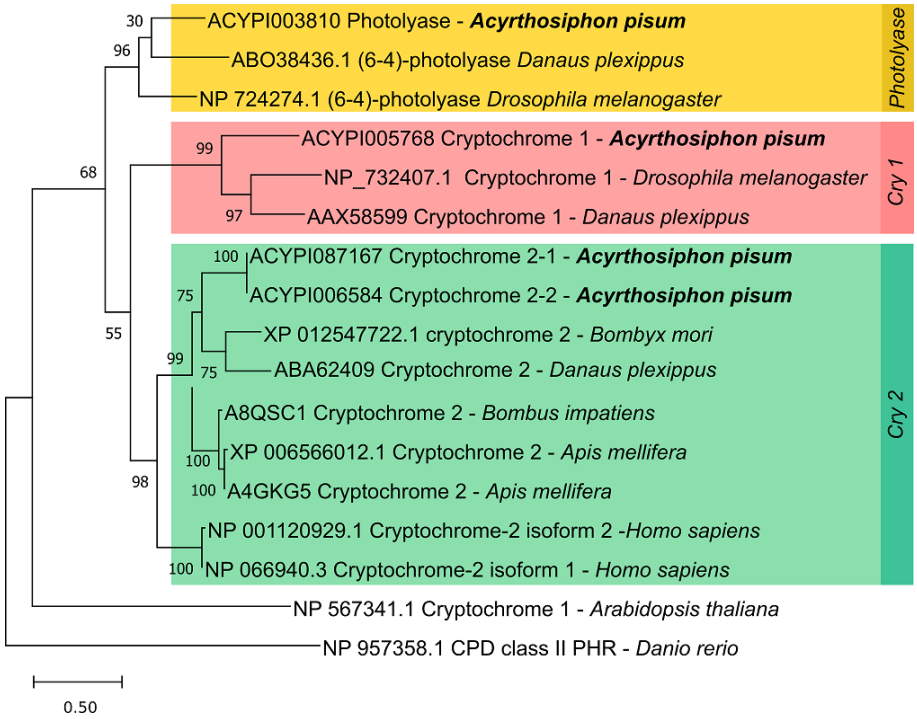


Figure 3.12: Insect cryptochromes and photolyase evolutionary analysis. Phylogenetic tree built using Maximum Likelihood method and JTT matrix-based model. Five hundred bootstrap pseudoreplicates were used. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Branch lengths represent the number of substitutions per site. Evolutionary analysis was conducted in MEGA X (Kumar et al., 2018). *Acyrthosiphon pisum* genes are shown in bold. Tags include accession number and species. *Danio rerio*'s CPD photolyase and *Arabidopsis thaliana*'s plant cryptochrome were used as outgroups (Mei et al., 2015). Cry 1: Cryptochrome 1 cluster (in pink); Cry 2: Cryptochrome 2 cluster (in green); (6-4) Photolyase cluster in yellow.

presence of those exons (a partial sequence alignment is shown in Figure 3.14).

From exons 5 to 12, no alternative splicing was found. The first 4 exons, however, showed some splicing variants (see Figures 3.13 and 3.14). This was already evident in the difference found between sequences NM_001171061.1 and FN377569.1, which vary in the distribution of the first 3 exons.

At least three alternative transcripts were found in this region (transcripts A to C). Transcript C is identical in exon structure to NM_001171061.1, which skips exon 3 and connects exon 2 with 4. Transcripts A and B keep exon 3. Transcript A matches exactly with FN377569.1, while transcript B lacks a 64 bp-long fragment that we called exon 3b (see Figure 3.13).

By observing amplicons in the agarose gel electrophoresis (Figure 3.15) some variants appear as minority products, as judged by their intensity in the stained gels, but the 3 splice variants are usually clearly detectable. As far as it is possible to determine by regular gel electrophoresis, variant C seems to be the most frequent transcript as revealed by the brighter bands found both in LSR1 and GR strain (see Figure 3.15).

Transcripts A and B differ by the inclusion or not of the exons 3 and 3b (see Figure 3.13). The presence or absence of the aforementioned exons didn't correlate with a particular aphid strain

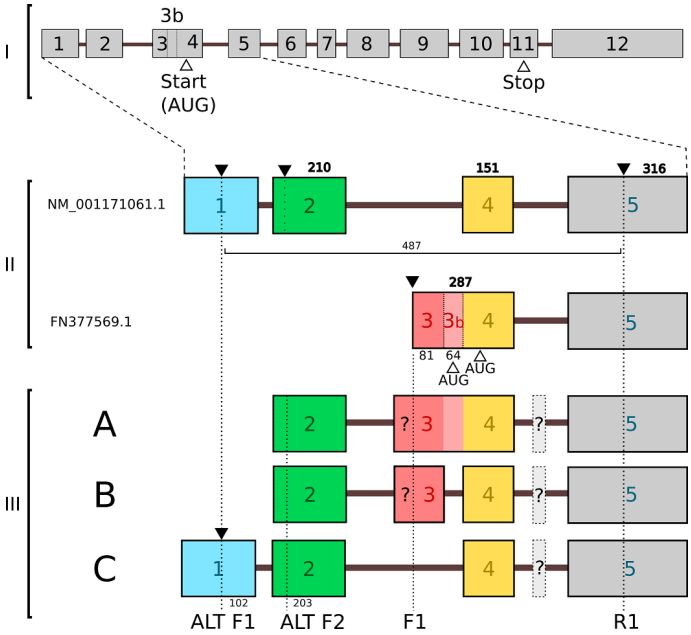


Figure 3.13: Structure of the *cry 1* gene and alternative transcripts found. **I)** All *cry 1* predicted exons in *A. pisum* combining the sequences available in the database. White arrowheads point at start and stop codons. **II)** Zoom into the first five exons in sequences NM_001171061.1 and FN377569.1 from the database. **III)** Alternative transcripts found experimentally. Exons are drawn to scale, numbered from 1 to 12. No alternative splicing was found in exons 5 to 12. A, B and C are proposed models of the 3 amplicons obtained through PCR (see text for details). Solid arrowheads: Primers AltF1, R1, AltF2 and F1 used in PCR (see Table 3.5). Hollow arowheads: AUG with open reading frames. Length of exons in nucleotides are shown over the exons. The uncertainty in the extension of exon 3 and the position of an putative extra fragment "X" (see the text) are shown by a question mark. Dotted lines show primer location.

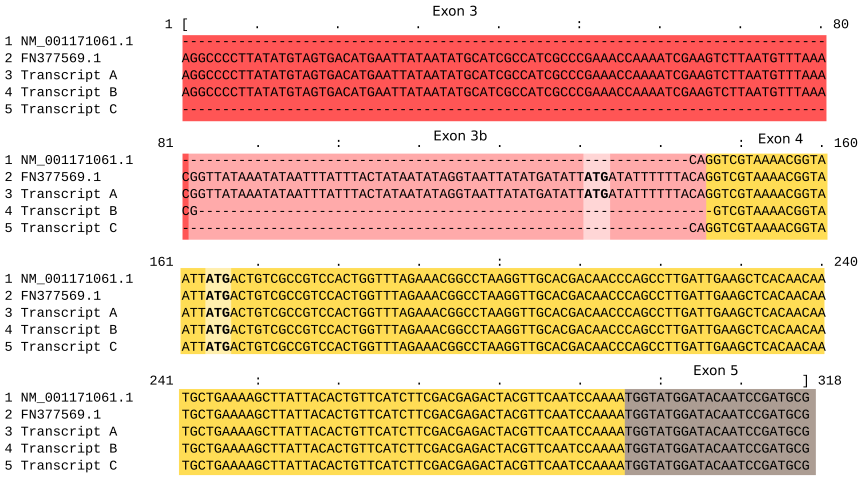


Figure 3.14: Partial sequence alignment of the first exons of *cryptochrome 1* alternative transcripts. Exons 3 to 5 (in the most variable region of the 5' region) are shown. Transcripts A and B have exon 3, while transcript C completely lacks exon 3 (as in the NM_001171061.1 sequence obtained from the AphidBase, http://bipaa.genouest.org/is/aphidbase/iagc/). Exon 3 in Transcript A has an extra fragment that we call 3b. Alternative starting codons are highlighted in exons 4 and 3b.

(GR or LSR1). One of the exons (exon 3b, Figure 3.13) presents an alternative AUG codon upstream of the predicted AUG in exon 4. Using this alternative starting codon would generate a protein 11 aminoacids longer, as it contains an open reading frame. The full length of exon 3 in its 5' region and its connection with exon 2 couldn't be sequenced.

A minoritary unpredicted fragment was also found. We called that amplicon "X", being ∼50 bp bigger than transcript A (Figure 3.16).

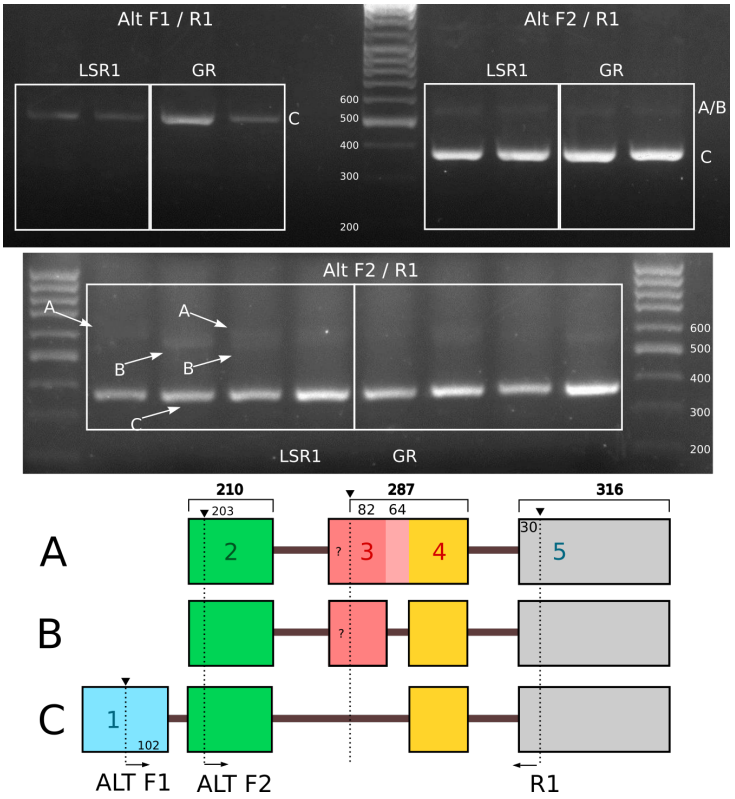


Figure 3.15: Agarose gel electrophoresis of *cry1* cDNA amplicons. The colored diagram below shows the interpretation of the results. A, B and C are the amplified transcript variants (see Figure 3.13). apCRY1-AltF1, apCRY1-R1 and apCRY1-AltF2 are primers used in the PCR. The prefix apCRY1- was omitted in the figure for simplification (see Table 3.5). DNA molecular weight marker in base pairs. Exons 6 to 12, which are unaffected by the alternative splicing, are not included in the diagram.

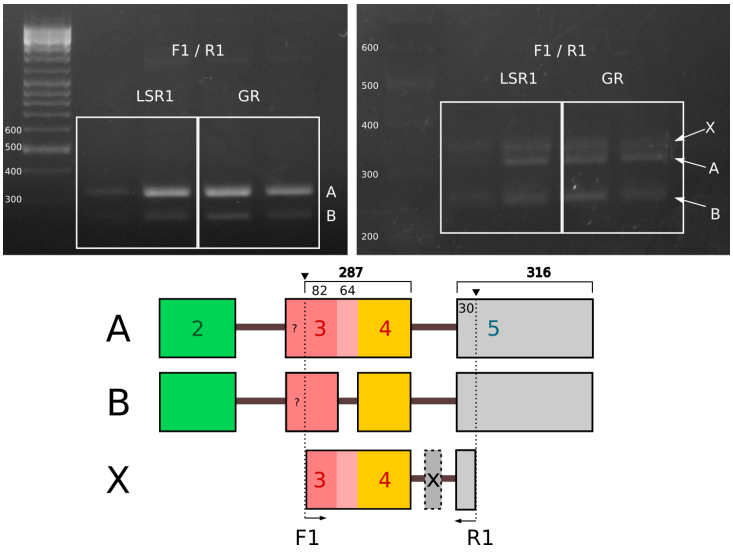


Figure 3.16: Agarose gel electrophoresis of *cry1* cDNA amplicons showing the presence of an unpredicted fragment. The colored diagram shows the interpretation of the results. A and B are the amplified transcript variants (see Figure 3.13). Variant X includes includes an extra fragment probably located between exons 4 and 5. apCRY1-F1 and apCRY1-R1 are primers used in the PCR. The prefix apCRY1- was omitted in the figure for simplification (see Table 3.5). DNA molecular weight marker in base pairs. Exons 6 to 12, which are unaffected by the alternative splicing, are not included in the diagram.

Analysis of *cry1* expression

We exposed both LSR1 and GR aphid strains to inducing (SD) and non-inducing (LD) conditions to compare the level of expression of the *cry1* gene. It was found that the holocyclic strain, LSR1, expressed slight but statistically significant higher levels of *cry1* than GR under both LD and SD photoperiodic conditions. Under LD conditions, LSR1 has 1.34 times higher expression than GR (1.29 vs 0.959) (Figure 3.17). Under SD conditions the difference between both strains is higher, with LSR1 showing 1.65 times higher expression when compared to GR (see Figure 3.17). Thus, the holocyclic strain LSR1 expressed higher levels of the *cry1* gene under both of the analysed photoperiods.

Globally, comparisons in strains didn't reveal differences between photoperiods: GR under SD conditions didn't show differences with GR at LD. Similarly, LSR1 exposed to SD didn't show higher expression levels than LSR1 under LD conditions.

To assess any difference between the first and second day of sampling, we grouped and compared the levels of expression of both days. Expression of *cry1* in the GR strain didn't show differences between the first and the second day.

LSR1 parthenogenetic females did reveal statistically significant differences on the second day, as it shows higher levels of expression

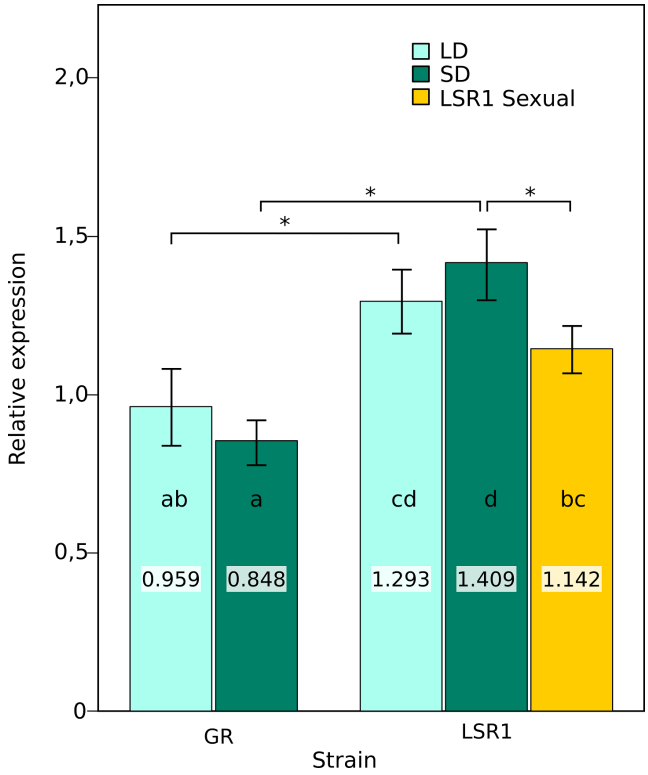


Figure 3.17: Relative expression of the *cry1* gene in two aphid strains under different photoperiods. LSR1, holocyclic strain; GR, anholocyclic strain; SD, Short Day conditions (12L:12D); LD, Long Day conditions (16L:8D). Error bars represent the standard error of the mean (SEM). Significant differences in gene expression are indicated by * ($p<0.05$)

under SD conditions when compared to LD conditions (Figure 3.18).

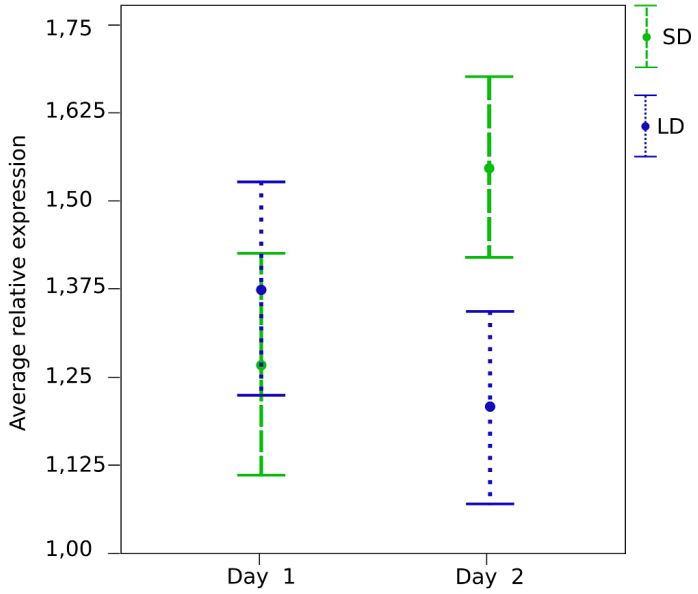


Figure 3.18: Average relative expression values of *cry1* in LSR1 parthenogenetic females in day 1 and day 2 under long day (LD) and short day (SD) photoperiods. Error bars: 95% CI.

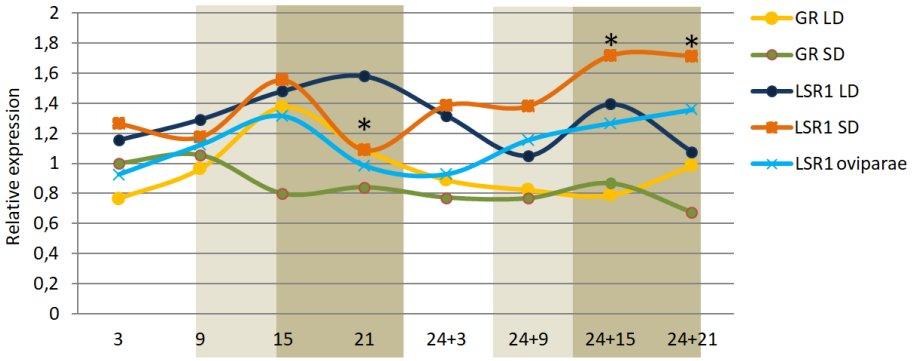


Figure 3.19: Relative expression of *cry1* for oviparae (sexual females, LSR1 strain) and parthenogenetic females (LSR1 and GR strains) of *A. pisum* at different time points during two consecutive days and under different photoperiods (see text for details). Asterisks show time points that display statistically different levels of expression than, at least, one of the other ZTs in the same strain and photoperiod (P<0,05). LSR1, holocyclic strain; GR, anholocyclic strain; ZT, zeitgeber time; SD, Short day; LD, Long day. Periods of darkness are shaded. Error bars have been omitted for clarity.

Analysis of circadian expression of *cry1*

With respect to a possible circadian rhythm in *cry1* expression, of all the conditions analyzed in this study, only strain LSR1 reared under SD photoperiod showed statistically significant differences between ZTs ($p<0.05$) (Figure 3.19). Both parthenogenetic and sexual females of this strain showed higher expression at ZT15 (3 hours after the onset of night) and decreased during the night at ZT21. Expression increased again during the day. Moreover, sexual females also showed a significant rhythm under COSINOR analysis ($p<0.05$), with an estimated period of 26 hours (Figure 3.20).

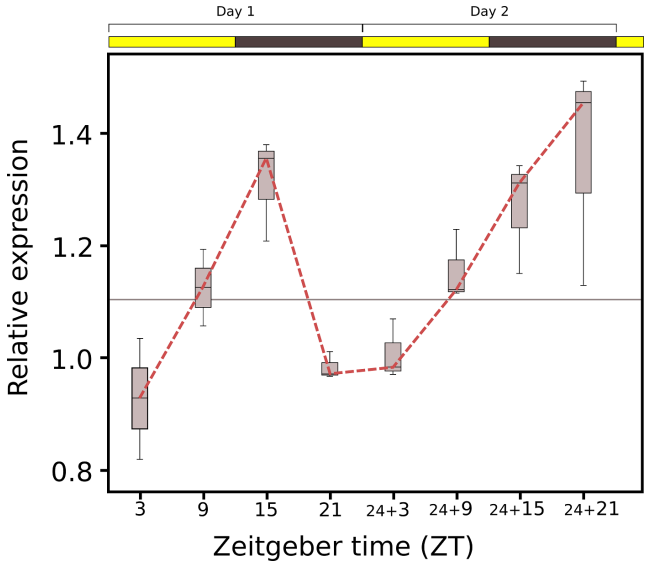


Figure 3.20: Daily relative expression of *cry1* in LSR1 sexual females under short day photoperiod (12 hr light : 12 hr dark). Periods of light and darkness are shown above the figure with yellow and black bars respectively. Horizontal line shows the rhythm-adjusted mean or MESOR (Midline Estimating Statistic Of Rhythm).

Transcript localization of *cry1* and *cry2*.

Cry1 riboprobes were bound to transcripts located in groups of neurons of the protocerebrum in the anterodorsal region, presenting a clear symmetrical mark of two cells in each hemisphere of the protocerebrum, overlapping or very close to the previously described clusters of small and large Dorsal Neurons (s-DN and l-DN) (Barberà et, al., 2017). Two more cells were symmetrically stained in the protocerebrum, one proximal to each optic lobe (Figures 3.21 and 3.24) overlapping with or proximal to the Lateral neurons (LN) (Barberà et, al., 2017).

Regarding *cry2*, the sites of expression were not as evident as in *cry1*. A weak signal was detected in the intersection between the protocerebrum and the optic lobe (see Figure 3.22 A), in the vicinity of the neurosecretory cells group V (NSC V), and the ventral and dorsal lateral neurons (vLN and dLN). Moreover, another region was detected (Figure 3.22) in the anterior region of the protocerebrum, probably in the *pars lateralis* (see Figure 3.22 B). Both signals are weak and the symmetry with the right portion of the brain can not be confirmed due to loss of that tissue during the dissection process or misplacement of the brain in the preparation.

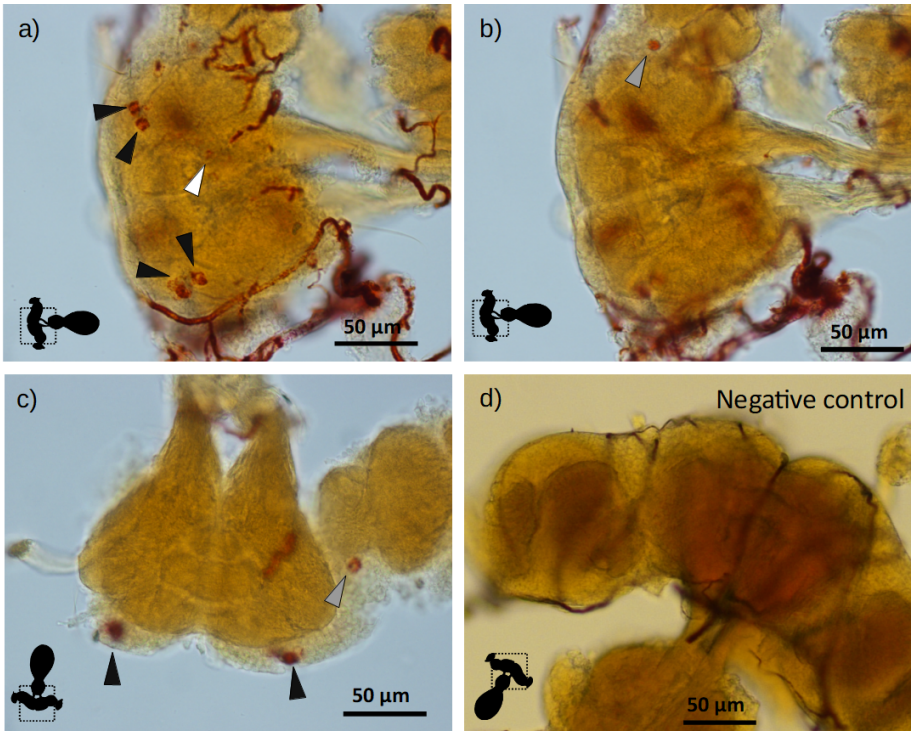


Figure 3.21: *In situ* localization of *cry1* gene transcripts in adult *A. pisum* central nervous system. Pictures a) and b) were obtained from the same preparation but with different focus adjustment. All arrows point to specifically stained regions near previously identified cells (see Figure 3.24). Black arrows point to cells close to dorsal neurons (DN); white arrows point to cells close to the neurosecretory cells of groups II and III (NSC II-III) and gray arrows point to cells close to dorsal and ventral lateral neurons (LN). A representation of the central nervous system is included in each image indicating the orientation of the preparation. d) The sense probe was used for negative control.

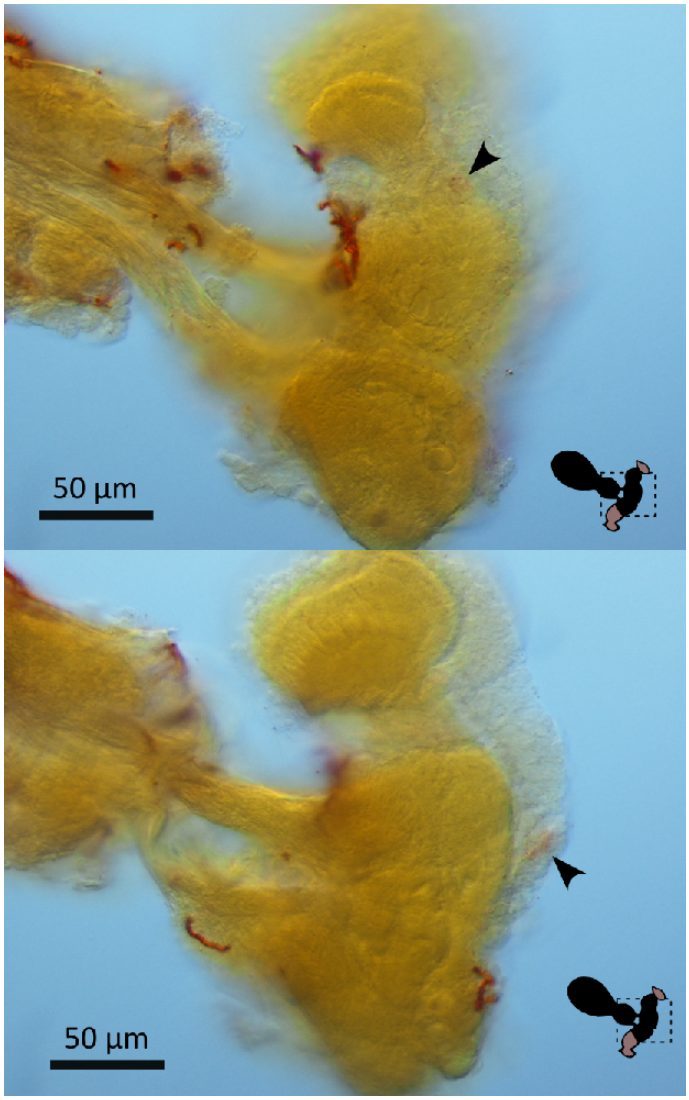


Figure 3.22: *In situ* localisation of *cry2* gene transcripts in adult *A. pisum* central nervous system. Arrowheads point to specifically stained regions. Top and bottom were obtained from the same preparation but with different focus adjustment. A schematic representation of the aphid central nervous system is included beside each image indicating the orientation of the preparation (in grey, areas of the brain or eyes lost during dissection process).

3.5 Discussion

In this chapter we have studied in detail the expression of two families of photosensitive proteins in the pea aphid. Members of both families, opsins and cryptochromes, are candidates to be photoperiodic photoreceptors. Previous studies such as those on light wavelength sensitivity in aphids (Lees, 1981) determined that medium wavelength (blue) light at particular moments during the dark period produced a long day effect in *M. viciae*. Cryptochromes and medium-wavelength-sensitive (MWS) opsins are sensitive to this blue light. Also, in the latter night, a light pulse of longer wavelengths (up to 600 nm, red light) could revert a SD response to LD. MWS and long-wavelength-sensitive (LWS) opsins could be involved in the photoreception process at this photosensitive moment of the night. We have studied opsin and cryptochrome daily patterns of expression and how they are affected by different photoperiods. Moreover, we have attempted to localize their specific site of expression in the aphid brain. All this information is valuable for discussing their putative implication in photoperiodic photosensitivity and in circadian rhythmicity.

3.5.1 Opsins

Aphid opsin gene repertoire

Although their spectral sensitivity was not experimentally determined, by looking at the phylogenetic position of the seven opsin genes identified (see Figure 3.4), the aphid opsin gene set includes at least three genes clustering with visual opsins, one of which (Ap-LWO) would be sensitive to long wavelength radiations and two (Ap-SWO2 and Ap-SWO3) would be sensitive to short wavelengths, as judged by their clustering in the phylogenetic tree. BLAST searches against both *A. pisum* and *M. persicae*'s genomes failed to retrieve any sequence similar to medium wavelength sensitive opsins (MWO), which are present in other insects (see Figure 3.4). Although MWO gene homologues are apparently absent in the *A. pisum* genome, studies on aphid spectral sensitivity suggest the presence of photoreceptors triggered by such light wavelengths in aphids (Döring et al., 2011; Kirchner et al., 2005). Thus, it is possible that aphids have gone through some spectral tuning of a different wavelength sensitive opsin in order to detect this medium wavelength radiations. Salcedo et al. (2003) used transgenic flies expressing mutant rhodopsins and found that the position corresponding to the G90 residue in bovine rhodopsin (K in invertebrates) is essential for UV spectral tuning. They show that the presence of a lysine

in that position shifts the absorption spectra into the UV range. It is noteworthy that Ap-SWO2 and Ap-SWO3 represent a recent duplication event of a short wavelength sensitive opsin gene. Both opsin genes conserve a similar intron pattern (see Figure 3.5) and are located only 8kb apart in the same genome scaffold (scaffold number GL350571). Most interestingly, only Ap-SWO3 presents a lysine in position G90, while Ap-SWO2 has a valine at that position (see Figure 3.6). It could be hypothesized that selection would have favoured a shift towards medium wavelength sensitivity of Ap-SWO2 absorption spectra. This evidence suggests a neo-functionalization of Ap-SWO2 opsin gene to account for a lack of a specific MWO pigment. The fact that the same sequence variation has been found in the corresponding *Myzus persicae* homologs (results not shown) further supports the view of a recruitment of a UV sensitive opsin for the detection of medium wavelength radiation in aphids (at least in members of the Aphidinae subfamily).

Apart from the three putative visual opsins, two aphid opsins (Ap-SWO1 and Ap-SWO4) clustered together with the *D. melanogaster* Rh7 protein. Although no much information was available until recently on this fly opsin partly due to its low expression levels in the tissues under study (Futahashi et al., 2015; Kistenpfennig, 2012) or for being "often overlooked" as stated by Feuda et al. (2016), a recent paper shows its involvement in entrainment of the fly circa-

dian clock by light (Ni et al., 2017). Kistenpfennig (2012) located Rh7 expression in antennal neurons and retina of *D. melanogaster* and suggested that it might be a shielding pigment for the highly expressed Rh1. We have detected Ap-SWO4 using PCR in brains and optic lobe/eye preparations and *in situ* hybridization also revealed mild expression in the protocerebrum and optic lobes (see Figures 3.2, 3.9 and 3.10). Interestingly, this signal was found in dorsal brain regions previously proposed to be involved in photoperiodism (Steel and Lees, 1977; Steel, 1978) that have been recently shown to contain clock neurons also expressing clock genes thus controlling circadian rhythms (Barberà et al., 2017), which would be in agreement with its role in circadian entrainment. However, further studies are needed to confirm co-localization with circadian and/or putative photoperiodic elements. Regarding Ap-SWO1, the second Rh7-like opsin found in aphids, we have been unable to detect its expression using riboprobes in aphid brains, but it has been the only opsin gene to be successfully PCR amplified in all antennal preparations in L3 aphids. This is consistent with previous findings of *D. melanogaster* Rh7 being expressed in the second antennal segment (Kistenpfennig, 2012) related to audition. The author, however, discarded a function of Rh7 in sound detection. Our findings suggest that a non-photosensitive role for this particular opsin should be considered. We also report on the presence in the aphid opsin

repertoire of a non-visual C-type opsin (Ap-C-Ops) that clusters with pteropsin, the arthropod C-type opsin found in bees, which has been proposed as a candidate clock photoreceptor expressed in the protocerebrum and optic lobes of *Apis mellifera* (Rubin et al., 2006). Most striking was the presence in the aphid opsin repertoire of a gene orthologous to arthropsin, an opsin gene that was first detected in the crustacean *Daphnia pulex*, where it has several duplications (Colbourne et al., 2011). Recently, arthropsin orthologues have also been reported in the spider *Cupiennius salei* and the velvet worm *Euperipatoides kanangrensis* (Eriksson et al., 2013). This opsin gene was apparently absent from insects until it was recently found in dragonflies (Futahashi et al., 2015). However, its presence in the pea aphid was already previously reported (Hering and Mayer, 2014), although the possible implications were not discussed in those studies. The present report confirms the presence of arthropsin in aphids but we consider that the distribution of this opsin gene in the phylogeny of insects deserves some more attention. Apart from its recently reported presence in Odonata (Futahashi et al., 2015), a BLAST survey (our unpublished data) confirmed the presence of arthropsin in the hemipterans *Diaphorina citri* and *Cimex lectularius*, which, considering also its presence in aphids, extends the presence of arthropsin to Sternorrhyncha and Heteroptera within Hemiptera. Additionally, if we consider the presence of arthropsin in Crustacea

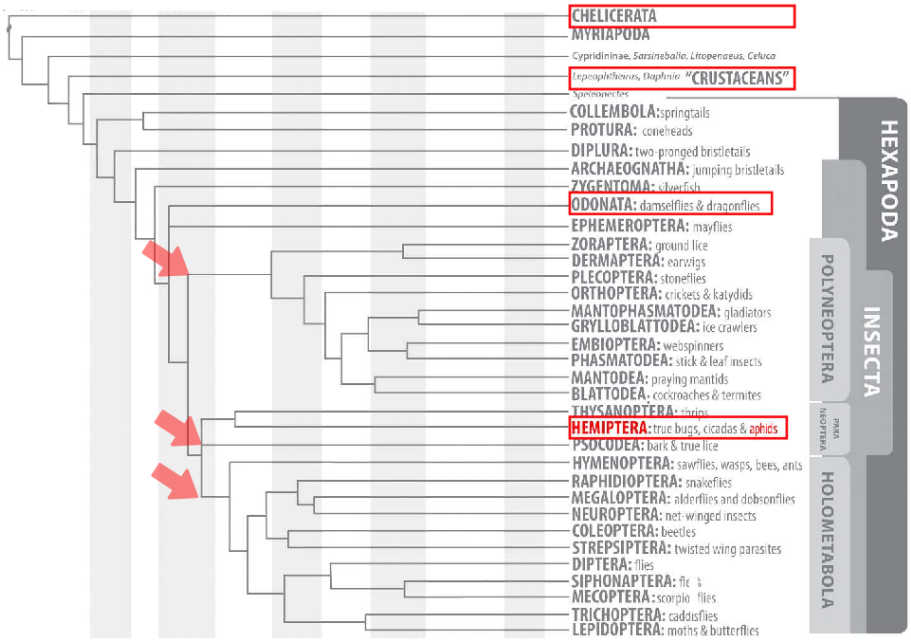


Figure 3.23: Distribution of Arthropsin gene in the insects group considering current information. The presence a Arthropsin genes in the crustacean *Daphnia pulex* highlights its ancestral origin. Independent losses in the ancestors Polyneoptera, Holometabola and Psocodea would explain this distribution (red arrows). Names of phyla where at least one member with an Arthropsin gene was found are in red boxes. Insect phylogeny modified from Kjer et al. (2016)

as well as in Chelicerata and Onychophora it seems reasonable to think that arthropsin represents an ancestral opsin and its current distribution within the class insecta implies independent losses in most insect groups (i.e., at least two or three independent loses in the ancestors of Polyneoptera, Holometabola and Psocodea, Figure 3.23).

Further research is needed to understand the nature of the selective pressures that have retained this opsin in aphids, while there seems to have been a tendency in most advanced insect groups sequenced to date to lose this gene. The presence of several copies of arthropsin genes in the genome of *Daphnia* is suggestive as this crustacean shares with aphids the cyclical parthenogenesis in which photoperiod also has a role inducing the sexual phase (Alekseev and Lampert, 2001). In this respect, it is interesting to recall that we have found a relationship between Ap-Arthropsin expression and the photoperiodic response in the pea aphid as holocyclic aphids exhibit increased expression of this gene under short photoperiod conditions. Moreover, the detection of arthropsin transcripts in the protocerebrum and the optical lobe, but not in the aphid eyes, suggests a relevant role in a non-visual photosensing process. The absence in 4 of the aphid proteins (Ap-SWO1, Ap-SWO4, Ap-C-Ops and Ap-Arthropsin) of the QAKKMNV amino acid motif highly conserved in the third intracellular loop in invertebrate opsins (Gao et al., 2000; Townson et al., 1998) is not completely unexpected as this motif is absent in other *D. melanogaster* Rh7-like proteins (Senthilan and Helfrich-Förster, 2016), vertebrate opsins and arthropsins. Ap-SWO1 and Ap-C-Ops both lack the N-glycosylation site widely conserved in opsins (See Figure 3.6). There are several potential roles for this post-translational modification in rhodopsins (Murray

et al., 2009), and according to the information available from *D. melanogaster*, this might indicate that protein maturation and transport could be affected (Katanosaka et al., 1998; O'Tousa, 1992). However, a role in the phototransduction cascade (if any) would still be possible according to O'Tousa (1992).

Opsin gene expression along the day-night cycle

Although the YR2 strain showed significant differences in levels of transcription of some opsin genes at particular ZTs, (i.e., Ap-SWO4 and Ap-Arthropsin both under SD and LD conditions and Ap-C-Ops under LD conditions, see Figure 3.7) the COSINOR analysis didn't reveal any significant rhythmicity. We consider, however, that further experiments are needed to provide additional data to definitely confirm or reject this apparent lack of rhythmicity, as we observe a cyclic trend in our data and other insects have proven to express opsin genes in a rhythmical manner, particularly long-wavelength and ultraviolet sensitive opsins (Claridge-Chang et al., 2001; Oba and Kainuma, 2009; Sasagawa et al., 2003; Wang et al., 2013; Yan et al., 2014). It is also possible that a background rhythm does exist in the expression of some opsin genes but with limited amplitude that prevents its discrimination from eventual background noise. Despite this lack of a significant rhythm, some trends could be observed in the daily expression of opsin genes (see

Figure 3.8). Expression of all opsin genes under LD conditions in the YR2 strain increased during the day until the evening at ZT9 and decreased thereafter until the next morning. Under SD conditions mRNA levels of four opsin genes (Ap-Arthropsin, Ap-LWO, Ap-SWO2 and Ap-SWO3) in the YR2 strain increased from the moment lights went on until a maximum in the latter half of the night (ZT21). Contrarily, Ap-SWO4 showed a minimum of expression at ZT21 its maximum peak being at ZT9 during daytime, similar to its expression under LD conditions (see Figure 3.8). Expression of opsin genes in the GR strain under LD conditions revealed a pattern very different to what was observed in the YR2 strain under the same photoperiodic conditions, with a minimum of expression at ZT9 (see Figure 3.8). GR strain exposed to LD peaked at ZT21 in most opsins, except for Ap-C-Ops and Ap-SWO4, whose maximum was observed at ZT15. Under SD conditions, the GR strain showed a peak of expression during the first half of the night (ZT15) for all opsins. The fact that most opsins shared a similar pattern of expression along the day-night cycle is striking but could be explained if they share a similar mode of regulation (e.g. the same transcription factors could control their expression) which could be possible given their common ancestry (especially for some recent duplications). There is variability between insect species in the daily pattern of expression of different opsins, being the long

wavelength sensitive the most studied opsin. While LWO does not seem to cycle in flies or mosquitoes (Claridge-Chang et al., 2001; Jenkins & Muskavitch, 2015), it decreases at dawn and shows higher expression levels at night in some coleopterans, hymenopterans and lepidopterans (Oba & Kainuma, 2009; Wang B. et al., 2013; Yan et al., 2014). This latter scenario seems to be the case in *A. pisum*. It has been proposed that increases in opsin expression at night or in dim-light environments improve vision in nocturnal insects, like moths. Aphids, however, are not as active and remain nearly motionless along the day, so an immediate benefit of an increase in visual pigments is less intuitive to deduce based on their behavior alone. However, an increase in light sensitivity at certain moments of the night could be in concordance with the External Coincidence Model, proposed to explain a relation between circadian clocks and photoperiodism (Bünning, 1960; Pittendrigh & Minis, 1964) (see General Introduction, section 1.3.1), with a photoinducible phase which, when illuminated, affects the photoperiodic response.

As mentioned in the General Introduction (section 1.3.1), the vetch aphid *M. viciae* has two sensitive moments: one early in the night and a second moment later in the night (Lees, 1973; 1981). In YR2 under SD conditions, in the latter night, at ZT21, Ap-LWO, Ap-Arthropsin, Ap-SWO3 and Ap-SWO4 had an expression peak. These genes might be in charge of photoperiodic photoreception in

that photosensitive phase of the night that reverts SD response to LD in Lees' experiments (1973; 1981).

Opsin gene expression in long days vs. short days

With respect to the quantification of opsin gene expression in relation with the photoperiod, short days (SD) seem to increase overall opsin expression. Increased opsin expression under SD conditions was also observed by Tamaki et al. (2013) in the opsin genes studied in the cricket *Modicogryllus siamensis*. In the aphid holocyclic strain YR2, opsin genes showed higher expression in aphids reared under SD than in aphids reared under LD conditions. However, in the anholocyclic strain (GR), only Ap-SWO1 and Ap-SWO4 under SD conditions were significantly overexpressed compared to LD. Ap-SWO1 showed strain specific behavior, with lower expression in the GR strain compared to YR2 regardless of the photoperiodic condition. By comparing levels of expression of the different opsin genes it looks like they show a basal level of expression in the anholocyclic GR strain irrespective of the photoperiod, which would be similar to the level observed in aphids of the holocyclic strain under LD (non-inducing conditions). Thus, while shortening of photoperiod seems to trigger an increase in expression of the entire opsin gene repertoire in the holocyclic strain, the anholocyclic strain does not modify levels of expression of most opsin genes in response

to photoperiod shortening. This result suggests a role for opsins in the photoperiodic response as anholocyclic aphids, in some way, could be considered blind to photoperiodic changes, which could be the consequence of their lost capability to trigger opsin gene expression activation by short photoperiods. However, since SD aphids of the holocyclic strain corresponded to adult sexuparae rather than to younger nymphal stages it is also possible that this increase of expression of opsin genes is not directly related with the signal causing the induction process but rather a consequence of this process. Still the problem remains of why should opsin genes have higher expression in shorter days than in long days. It is possible that these proteins might increase their level of expression to optimize light sensitivity and/or temperature sensing, as it has been reported that these GPCRs are also involved in temperature discrimination in *D. melanogaster* (Shen et al., 2011). In this respect it must be regarded that temperature acts secondarily along with day length in the activation of the photoperiodic response in aphids, being responsible for fine tuning the photoperiodic response.

It is worth mentioning that recent findings by our group (Barberà et. al, 2019) have revealed that the expression of ILP1 and ILP4, two Insulin-Like Peptides in *A. pisum* is opposite to that observed here for opsins: it is repressed under SD conditions in the holocyclic strain. As mentioned in section 1.3.1, our group proposed

ILPs to be the so called "virginoparin", the signal that induces virginoparae production (Steel, 1976). The higher opsin levels under SD conditions might be the starting point of a pathway leading to ILP inhibition, thus allowing sexual morphs production.

Localization of opsin gene expression

Localization of Ap-SWO2, Ap-SWO3 and Ap-LWO in the aphid eyes both by PCR on dissected head parts and by *in situ* hybridization (see Figures 3.9 and 3.10) is in concordance with the information obtained from sequence and phylogenetic analysis, which convey a visual role to these proteins. The intense Ap-LWO signal detected in eyes was also seen in its orthologues in other insect species, like the orthopteran *Modicogryllus siamensis* (Tamaki et al., 2013) and the lepidopteran *Danaus plexippus* (Briscoe, 2008). This is also consistent with the observations in *D. melanogaster*'s long-wavelength sensitive opsin, Rh1, which is expressed in six out of eight photoreceptors cells in every ommatidium (O'Tousa et al., 1985). Similarly, the dispersed expression observed for Ap-SWO2 and Ap-SWO3 (Figure 3.10) reminds of *D. melanogaster*'s retina, with short-wavelength sensitive homologs (Rh3 and Rh4) being expressed in one or two cells per ommatidium (Cook and Desplan, 2001), which generates a more scattered distribution of these proteins in the eye. We might discard a role for these three visual opsins as main

input elements for the photoperiodic response, since early studies on the aphid photoperiodic response demonstrated that photic detection would be located in the dorsal region of the protocerebrum rather than in the compound eyes (Lees, 1964; Steel and Lees, 1977). We cannot discard, however, that these receptors could provide auxiliary information on prevalent photoperiodic conditions to provide robustness to the main (yet unknown) input pathway.

Aphid opsin homologs which do not cluster with insect visual opsins (Ap-Arthropsin, Ap-SWO1, Ap-SWO4 and Ap-C-Ops) were PCR-amplified not only in OL/eyes preparations, but also in protocerebrum and ganglia sections (see Figures 3.2 and 3.9). *In situ* hybridization confirmed and provided further information on a more precise localization of these genes' expression by detecting Ap-Arthropsin, Ap-SWO4 and Ap-C-Ops transcripts in the optic lobe and the protocerebrum (Figures 3.10 and 3.24). We were unable, however, to confirm opsin expression in ganglia via *in situ* hybridization due to difficulties eliminating background staining caused by abundant trachea in aphid preparations in that region. As for the role of the optic lobes in photoperiodic response, Steel & Lees (1977) regarded them as "inessential for the photoperiodic mechanism". Thus expression of Ap-Arthropsin, Ap-SWO4 and Ap-C-Ops in the optic lobe could be related with other (unknown) purposes or be involved in light-independent processes, such as thermosensitivity

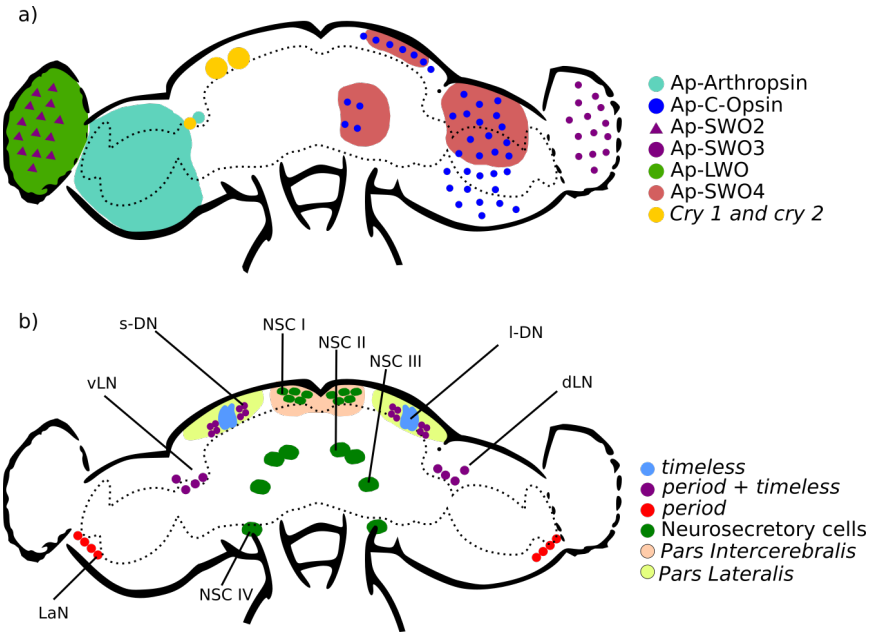


Figure 3.24: Schematic localization of the different opsin and cryptochrome genes expression in the aphid brain. **a)** Opsin and cryptochrome transcripts localization. For simplicity, expression is depicted only in one half, but was found symmetrically in both halves of the brain. **b)** Localization of relevant brain structures, *Pars lateralis* and *Pars intercerebralis*, neurons expressing clock genes *timeless* and *period* and neurosecretory cells. LWO, Long Wavelength Opsin; SWO, Short Wavelength opsin; NSC, neurosecretory cell; DN, dorsal neuron (s- small, l- large); LaN, lateral neuron; dLN and vLN dorsal and ventral lateral neurons. Adapted from Barberà et al. (2017).

(Shen et al., 2011) or even both. Regarding opsin expression in the protocerebrum, it is noteworthy that both Ap-SWO4 and Ap-C-Ops transcripts were also detected in the dorsal region known as *pars lateralis* (Figures 3.10 and 3.24). This region was found relevant to prevent photoperiodic induction when exposed to light (Lees, 1964), which leaves Ap-SWO4 and Ap-C-Ops as good candidates to participate in the photoperiodic input pathway. Ap-C-ops and Ap-Arthropsin were also detected close to lateral neurons in the protocerebrum, where the clock related genes *period* and *timeless* were found to be expressed (Barberà et al., 2017), so they might as well be involved in photosensing in circadian rhythms and/or photoperiodic photosensitivity.

Figure 3.24 schematically summarizes the localization of the expression of the different opsin genes and compares it with the localization of clock expressing neurons.

3.5.2 Cryptochromes

In this thesis we have revised the structure of *cry1* and analyzed the expression of the *cry1* and *cry2* genes in *A. pisum* in the context of photoperiodism in this aphid species. We PCR amplified *cry1* transcripts and quantified its levels of expression along the day/night cycle under two different photoperiod conditions both in a holocyclic and an anholocyclic aphid strain. In addition, we identified the site of expression of *cry1* and *cry2* in the aphid central nervous system. The information obtained was analyzed and compared with the available data (mainly from *D. melanogaster*) to asses a possible role of these genes in the photoperiodic response either directly or indirectly through the circadian rhythms in aphids. Regardless of the similarities in structure and sequence between the elements that make the clocks of *D. melanogaster* and *A. pisum*, it is expected for the fly and aphid circadian clockworks to show important differences in their functioning. First, the aphid has a mammalian-type cryptochrome gene (*cry2*) which is absent in *D. melanogaster* (Cortés et al., 2010). This means that the aphid clock must be more similar to an ancestral insect system also present in other insect groups, like butterflies (Yuan et al., 2007). Second, aphids don't have a *jetlag* homologous gene, which, at least in *D. melanogaster*, has been shown to be involved in cryptochrome-

mediated degradation of TIM (which is essential for clock resetting and thus for its synchronization with the day-night cycle).Thus, how aphids manage to degrade TIM without JET remains unknown and therefore it is not known how they synchronize their daily rhythms with the day/night rhythm. Moreover, it is even unknown if CRY1 has any role in clock reseting through TIM degradation in aphids.

Transcript size variability in *cry1*

Our results suggest that *cry1* goes through post transcriptional regulation through differential splicing. Indeed, we have detected 4 different amplicons in the 5' region of the transcripts corresponding to the presence or absence of alternative exons (transcripts A, B, C and X) (see Figures 3.13, 3.15 and 3.16). Although the relative abundance of some of these transcripts is low (as revealed by the faint bands in the agarose gels of their corresponding amplicons) and we couldn't determine a predictable pattern of transcript variants, they were systematically amplified on many aphid samples both from the LSR1 and the GR strains. Thus we can not relate their presence or absence with the condition of being holo or anholocyclic. The variant we refer to as "C" seems to be the majority transcript (see Figure 3.13).

Some minor details regarding splicing variants remain to be fine tuned, as the precise extension of exon 3 in its 5' region or the

source of the minority fragment between exons 4 and 5 that we call "X" (see Figure 3.16).

Further experiments are needed to determine if there is any difference in the level of expression of each variant in the different strains or under different photoperiods. A quantitative difference could have a biological meaning for the role of different alternative *cry1* transcripts in the process of aphid response to photoperiodism.

Differences in expression between photoperiods and strains in *cry1*

RT-qPCR results showed that the holocyclic strain, LSR1, expresses higher levels of *cry1* than GR under LD and SD photoperiodic conditions (1.34 and 1.65 times higher levels than GR, respectively). This rise in transcription levels most probably increases the availability of CRY1 protein and could have an effect in the sensibility of the aphid to the shortening of the photoperiod if it is effectively involved in the process. In *D. melanogaster*, the light regime doesn't have a strong effect on *cry1* transcript regulation, although light does regulate the levels of CRY1 protein in the fly (Emery et al., 1998). This result seemed to differ with the results obtained by Cortés et al. (2010) for the pea aphid, where strain LSR1 presented higher expression of *cry1* under SD conditions. Although our experiment also shows higher expression of *cry1* in LSR1 under SD conditions,

the difference is not statistically significant. This difference, however, could be due to the longer sampling time in the current experiment, which covers two days instead of one, as was the case in Cortés' et al. (2010) work. To ascertain this, we analyzed differences in the level of expression grouping the values obtained in the first and second day of sampling in LSR1 parthenogenetic females (Figure 3.18). In day 1, there are no differences between LD and SD aphids. In day 2, however, expression under SD conditions increases, being significantly higher than in day 1 and also than LD aphids of the second day ($P<0.01$). This shows a local difference that might have been detected in the experiment by Cortés et al. (2010). Also, it reveals that *cry1* expression under SD conditions increases with time, which might be connected to processes related with aphid development, as aphids in day 1 had recently gone through molting (see Materials and Methods in section 3.4.2). *cry1* expression seems to be strain specific, as LSR1 had significant higher levels of *cry1* mRNA than GR. It remains unknown whether this is directly related with the incapability of anholocyclic strains to respond to photoperiod shortening. It is possible that lower levels of *cry1* expression in the anholocyclic strain derive in lower protein levels, which could affect processes occurring downstream of *cry1* mediated regulation, such as differential sexual morph production. It is of great interest to analyze other strains, both holocyclic and anholocyclic, so as to find

out if their responsiveness to photoperiod (or lack of it) involves the same molecular elements as LSR1 and GR or if, on the contrary, the variability in photoperiodic response was achieved through different mechanisms.

Lack of rhythm of *cry1* in parthenogenetic females

Although *D. melanogaster* does show clock controlled rhythmic levels of *cry1* mRNA, research on *Drosophila*-like cryptochrome in other species reveal that other arthropods don't (Gentile et al., 2009; Meuti et al., 2015; Rund et al., 2016; Zantke et al., 2013), including the LSR1 strain of *A. pisum* (Cortés et al., 2010). We confirm those results previously reported in parthenogenetic aphids and extend the lack of rhythmicity to the anholocyclic strain (GR) under the two photoperiods studied. However, this observation doesn't discard a possible post-translational daily variation of the CRY1 protein in aphids, as is the case of *D. melanogaster* CRY1, whose protein levels also vary rhythmically. In the case of the fly, CRY1 protein is regulated by light at the translational or post-translational levels (Emery et al., 1998; Ozturk et al., 2013).

Rhythmic transcription of *cry1* in LSR1 sexual SD females

Differently from what was observed in ordinary parthenogenetic females, significant differences were found in the expression of the

cry1 gene between ZTs in LSR1 aphids under SD conditions, both in sexuparae and in sexual females. However, only the latter also showed a significant circadian rhythm after a COSINOR analysis ($p<0.05$) (see Figure 3.20). Although *D. melanogaster* also shows rhythmic expression of *cry*, it presents a very different pattern of mRNA levels, with peaks at ZTs from 1 to 5 and a trough from ZT17 to ZT19 (Emery et al., 1998). In the aphid sexual females we found almost the opposite, where maximum expression levels are reached around ZT15, and minimum levels at ZT3, meaning that aphid protein levels might increase in the latter night (see Figure 3.20). However, the protein expression values of CRY1 in *Drosphila* is very different from that of its mRNA. Due to a strong translational or posttranslational regulation, *Drosophila* CRY1 protein levels increase during the day and reach a maximum at night (Emery et al., 1998), thus showing a pattern more similar to that of *A. pisum*'s *cry1* mRNA. Whether fly and aphid CRY1 protein levels accumulate at the same time of the day remains unknown until data for CRY1 protein from aphids is available. There is a shift of about five hours between peak levels of the aphid mRNA and fly CRY, but that difference could be shortened by a delay in translation.

Localization of *cry1* and *cry2* expression in aphid brains.

In the present report, we have been able to clearly locate *cry1* mRNAs and to find signals of *cry2* in two brain regions close to the recently described clock neurons in aphids, i.e., in the dorsal and lateral neurons clusters in the protocerebrum where *per* and *tim* clock genes are expressed (Barberà et al., 2017) (see Figures 3.21 and 3.24). These two clusters have been described as relevant to the circadian clock in several organisms, including *A. pisum* (Barberà et al., 2017). First, *cry1* was detected close to small and large Dorsal Neurons (s-DN and l-DN, respectively) in the *pars lateralis* of the protocerebrum, which express, at least, two key clock genes: *timeless* and *period* (Barberà et al., 2017). Interestingly, this is the region where other insect species also harbor the central circadian machinery and, moreover, it corresponds to the region where *D. melanogaster* expresses *cry1* (Emery et al., 2000). Additionally, in the classical experiments of microcauterization performed by Steel and Lees (1977), damaged induced to the *pars lateralis* (close to the *pars intercerebralis*) prevented the production of parthenogenetic females under LD conditions (see Figure 1.20). Thus *cry1* is found a in region classically known for being relevant to the photoperiodic response and where essential clock neurons are located. Also in this region, close to the *pars lateralis*, *cry2* ribroprobes showed a weak

signal that needs further experiments to be confirmed (see Figure 3.22). The CRY2 protein in other organisms is not photosensitive, thus its presence in this region could not be related to photosensitivity but to a regulatory role as transcription factor for clock elements (probably in collaboration with PER and/or TIM, which are also expressed in those cells).

The second region were *cry1* was found is close to Lateral Neurons (LN), where *timeless* and *period* have also been detected (Barberà et al., 2017). Unfortunately, microcauterization experiments in this region of the brain, which includes the group of Neurosecretory Cells V, were not conclusive in their findings (Steel & Lees, 1977). Further analyses is needed to determine if they are closer to ventral or dorsal lateral neurons (vLN and dLN respectively), but their proximity to this region where *timeless* and *period* are also expressed is worth being noted. Again, *cry2* preparations showed signs of expression in this region of the brain (see Figure 3.22). Although the staining is not as manifest as in other genes in this work (like *cry1* or the opsins), a signal in the region known to express clock genes might result relevant. We discard the possibility of cross-reaction between riboprobes for *cry1* and *cry2* as they were synthesized from non-shared regions of their nucleotide sequences (recall that they are paralogous genes).

Despite that colocalization is not strictly confirmed by this

experiment, these findings place CRY1 as a strong candidate to act as a photoreceptor relevant to the input pathway of the circadian clock and/or the photoperiodic response (see Figure 3.24). Moreover, the partial results found for *cry2*, although not conclusive, are in accordance with a role also for CRY2 in the circadian and/or photoperiodic molecular machinery, likely not as a photoreceptor, but as a transcriptor factor.

Chapter 4

General discussion and conclusions

4.1 General discussion

The molecular mechanisms that rule the aphid photoperiodic response are, probably, highly complex. The current body of knowledge is built from pieces of information obtained through different techniques and analysing diverse aspects of this phenomenon. Much of that information is not connected to a general view and many gaps remain unfilled. In this work, we made use of transcriptomics to obtain a broader view of the phenomenon to help us connect isolated elements and integrate them into a general view. On the other hand, other investigations in our group have studied the puta-

tive output and core elements of photoperiodic and circadian clocks. This work aims to provide specific information on the remaining element of the photoperiodic response: the input pathway. We have done this by characterizing the most important candidates to be photoreceptors in the process: opsins and cryptochromes.

Through a comparative transcriptional analysis of three different photoperiods (one long and two short), we found that aphids exposed to shorter photoperiods present a higher number of differentially expressed genes when compared with aphids reared under long photoperiods. This is probably caused by aphids being exposed to more extreme conditions. There are, however, many pathways that are affected in the two short photoperiods analyzed. These common pathways are probably involved in a general response to short photoperiods, be it male or oviparae-producing photoperiod. We found that both of these two different outcomes of shortening of the photoperiod make use of hormones, neuropeptides and neurotransmitter metabolism. Levels of serotonin, melatonin, juvenile hormone, dopamine and insulin-like peptides are affected when the days shorten, signaling the arrival of the harsh winter season. Our evidence supports the existence of a common pathway that detects short photoperiods but ends up producing either males or oviparae depending on the duration of light availability. Among the differences that could be involved in deciding which sexual

morph to produce, we found the differential expression of *takeout* or *takeout*-like genes, that could be linked to sex-biased progeny production.

The correct measuring of the light input is, of course, of vital importance to aphids, as the production of the sexual phase (i. e. of males and sexual females) ensures the production of overwintering eggs. Regardig light sensitive proteins, we have shown that the whole opsin repertoire in the pea aphid is affected by the photoperiodic conditions, as increased expression of all opsin genes in a holocyclic strain was observed when reared under short photoperiodic conditions. In addition, three of these opsins (Ap-SWO4, Ap-C-Ops and Ap-Arthropsin) might be also involved in aphid circadian rhythms, given the localization of their transcripts close to dorsal and lateral neurons in the protocerebrum, where the clock genes *period* and *timeless* have been shown to be expressed (Barberà et al., 2017). Expression of opsin genes tends to be higher at night, probably increasing sensitivity during a photoinducible phase to elicit the photoperiodic response, which would be in accordance with the External Coincidence Model, also known as Bünning's Hypothesis (Bünning, 1960). We propose a visual function for Ap-LWO, Ap-SWO2 and Ap-SWO3, with Ap-SWO2 being a UV sensitive duplicate likely tuned to be able to detect medium wavelength (blue) radiations. The aphid C-opsin (Ap-C-Opsin) and the aphid

Rh7 homologue (Ap-SWO4) emerge as good candidates to act as photoperiodic receptors, as they are located in anterodorsal brain regions close to neurosecretory cells known to be necessary for the photoperiodic response (Steel and Lees, 1977). Ap-Arthropsin is an ancient opsin present in hemipterans but absent in other insect lineages (e. g. in Holometabola). Ap-SWO1 and Ap-SWO4, which cluster together with the poorly characterized *Drosophila* Rh7-like opsins, also show expression levels affected by the photoperiod. Ap-SWO1 is the only opsin to have been unequivocally amplified via PCR in antennas. Given the increasing amount of different functions in which opsins have been involved, further studies are needed to precisely describe the role of every opsin present in the aphid gene repertoire and to better understand their role in the photoperiodic response.

The information obtained in this study is consistent with the participation of *A. pisum*'s *cryptochrome* genes in photoreception for photoperiodism. Aphid *cry1* and (most likely) *cry2*, are expressed in brain regions related to the circadian clock. Moreover, *cry1* is expressed in a rhythmic manner in sexual females. The holocyclic strain expresses higher levels of *cry1* than the anholocyclic one, which might be related with the main difference between these strains: the ability to produce or not a photoperiodic response. Whether this strain-specific variability in *cry1* expression is a cause

or consequence of the difference in the life cycles of these morphs has yet to be solved.

This work was done as part of a greater project carried out by our group aiming to understand the mechanisms ruling the photoperiodic response in aphids. The bigger project studies not only the input mechanisms studied in this report, but also the putative core and output pathways of the photoperiodic clock as we try to fill in the gaps of a complex landscape still to be fully elucidated. Therefore, we will describe how the information obtained in this study (Figure 4.1) could fill some of the gaps on the proposed molecular mechanism for aphid photoperiodic response depicted in Figure 1.21 in the the General Introduction (section 1.4).

Regarding the photoperiod response, *A. pisum* likely perceives the presence or absence of sunlight in the *pars lateralis*. We found three genes coding for photosensitive molecules to be expressed in that region, close to the Dorsal Neurons that express *tim* and *per*: Ap-SWO4, Ap-C-Opsin and *cry1* (see sections 3.5.1 and 3.5.2). We also found overexpressed genes and pathways related to synthesis of opsins and to phototransduction pathways involving opsins in the aphid brain under different photoperiods (see section 2.5). Thus, these opsins and *cry1* make good candidates for photoperiodic photoreception. Also in the *pars lateralis* we detected *cry2*, where clock genes *timeless* and *period* are expressed. This other (and probably

not photosensitive) cryptochrome is most likely playing a role as a regulator element in the core of both the photoperiodic and, probably, the circadian clock. The core of the photoperiodic clock would keep track of day length and count the number of successive short or long days. This information would then be transmitted to the *pars intercerebralis* where the group I of neurosecretory cells (NSC I) is found. The NSC I is necessary for photoperiodic response and there is where some insulin-like peptides (ILPs, the putative "virginoparin") are synthesized. In this study we have found differential expression of genes and pathways involving synthesis and regulation of ILPs and insulin receptors. We have also found DE genes related with the metabolism of melatonin, serotonin (its precursor) and dopamine (melatonin and dopamine are thought to be mutually inhibited) (see section 2.5). We have also found DE genes involved in juvenile hormone synthesis and regulation and in FoxO signaling pathway, both of which would act downstream of the ILPs and affect diapause and development in other insects (Sim et. al, 2013; 2015) (see section 2.5). An increase in ILPs (most probably ILP1 and ILP4) would be sensed as a "Long Day" signal and via direct innervation with ovaries and embryos would promote virginoparae instead of sexual embryo development. That innervation might be mediated by the DE brain remodelling pathways we found in this study. Depending on the perceived length of the photoperiod, the

next generation will consist mainly of males (SD_{14} conditions) or sexual females (SD_{10}). In this regard, as almost all the enriched pathways found under SD_{14} are also present under SD_{10}, there must be a common pathway for both sexual outcomes, but slight differences lead to greater proportion of males or sexual females. Regarding those differences, we have found three *takeout*-like genes underexpressed in male producing conditions and two of them overexpressed in oviparae producing conditions. The *takeout* gene is, thus, a good candidate for driving sex determination in aphids.

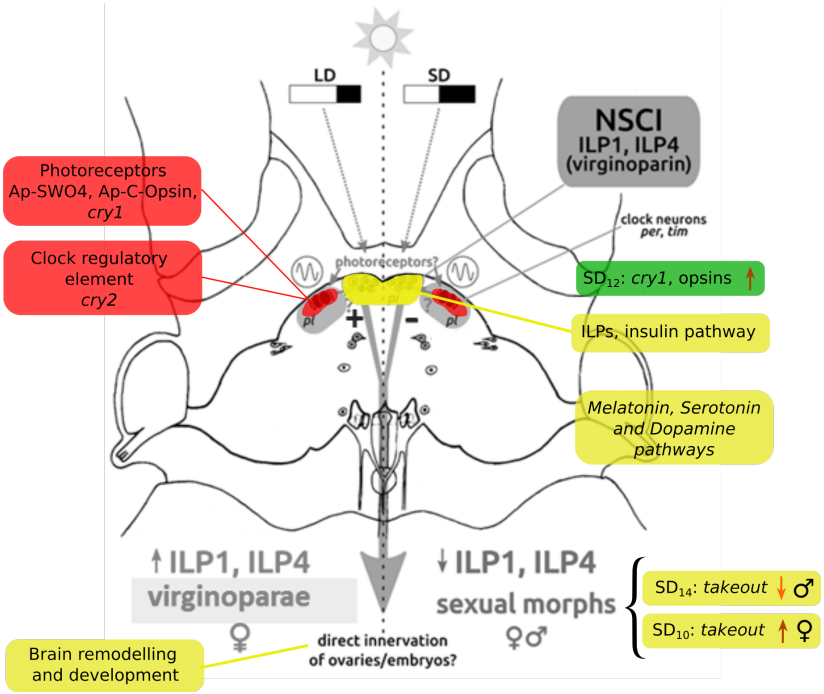


Figure 4.1: Contributions of this study to the identification of molecular elements involved in the aphid photoperiodic response. See Figure 1.21 for reference. In color, information obtained in this study. Yellow: transcriptomic results; red: *in situ* hibridizations; green: RT-qPCR results (see text). Long days (LD, left half of the graph) or short days (SD, right half) would be perceived through photoreceptors, such as Ap-SWO4, Ap-C-Opsin or *cry1* in the *pars lateralis* (*pl*) region, most probably in the Dorsal Neurons, where the relevant core clock genes *per* and *tim* are also expressed. In the *pl*, *cry2* could act as a clock regulatory element. The message would then be sent to the *pars intercerebralis* (*pi*) through and unknown mechanism, where the group I of neurosecretory cells (NSCI) expresses Insulin-like peptides 1 and 4 (ILPs). ILP 1 and 4 have been proposed to reach embryos and affect development. Melatonin, juvenile hormone and dopamine are strong candidate effectors in the photoperiodic output. Differential expression of *takeout* or *takeout*-like genes could promote either male or sexual female embryo development.

4.2 Conclusions

1. The exposure of aphids to different regimes of short photoperiods produces differential expression of genes related to hormone metabolism, neuropeptides and neurotransmitter biosynthesis (such as serotonin, JH, dopamine and melatonin) and insulin-like peptides, which are well known participants in the photoperiodic response in other insects.

2. Consistent with a response induced by light availability, our transcriptomic study found enrichment of pathways responsible for light perception, such as those needed for opsin synthesis and phototransduction.

3. Our results show that photoperiod has an impact on brain functioning, as revealed by the enrichment of pathways related to brain remodelling, increased transcription and gene regulation in the brain.

4. Digestive enzymes were found to be affected by the length of the photoperiod, which might be related to a putative effect of the light regime on the plants where aphids feed.

5. *takeout*-like genes seem to be related to sex-biased progeny production. They are underexpressed in SD_{14} and overexpressed in SD_{10}.

6. Many (>40%) of the top DE genes in the photoperiodic induction of sexual aphid morphs are uncharacterized transcripts. Further analysis is needed to understand their role.

7. An opsin repertoire of 7 genes has been identified in the aphid genome.

8. Genes belonging to the photosensitive protein family of opsins are overexpressed under SD photoperiods.

9. The photosensitive protein *cry1* shows higher expression in the holocyclic strain LSR1 than in GR. LSR1 strain also overexpresses *cry1* under SD conditions.

10. Opsins Ap-SWO4 and Ap-C-Ops are expressed in (or very close to) the dorsal neurons (DN) of the *Pars lateralis* shown to be relevant to photoperiodic response, where the clock genes *tim* and *per* are expressed. Also, these two opsins together with Ap-Arthropsin were found in the vicinity of the lateral neurons (LN) also expressing *tim* and *per*.

11. *cry1* (and likely *cry 2*) is expressed in the *Pars lateralis* close to Dorsal and Lateral Neurons (DN and LN), where *timeless* and *period* genes are expressed.

12. *cry1* is expressed in a circadian manner in *A. pisum* sexual females.

13. We hypothesize that the lack of a medium wavelength sensitive opsin in *A. pisum* can be compensated by the lysin to valine mutation in position G90 of Ap-SWO2 that could favor a shift towards the medium wavelength absorption range.

Bibliography

Adams, J., Kelso, R., & Cooley, L. (2000). The kelch repeat superfamily of proteins: Propellers of cell function. *Trends in Cell Biology*, *10*(1), 17–24. doi:10.1016/s0962-8924(99)01673-6

Ahmad, S. T., Joyce, M. V., Boggess, B., & O'Tousa, J. E. (2006). The role of *Drosophila* ninaG oxidoreductase in visual pigment chromophore biogenesis. *Journal of Biological Chemistry*, *281*(14), 9205–9209. doi:10.1074/jbc.m510293200

Alekseev, V. & Lampert, W. (2001). Maternal control of resting-egg production in *Daphnia*. *Nature*, *414*(6866), 899–901. doi:10.1038/414899a

Allada, R., White, N. E., So, W., Hall, J. C., & Rosbash, M. (1998). A mutant drosophila homolog of mammalian clock disrupts circadian rhythms and transcription of period and timeless. *Cell*, *93*(5), 791–804. doi:10.1016/s0092-8674(00)81440-3

Apfeld, J. & Kenyon, C. (1998). Cell nonautonomy of *C. elegans daf-2* function in the regulation of diapause and life span. *Cell*, *95*(2), 199–210. doi:10.1016/S0092-8674(00)81751-1

Arquier, N., Géminard, C., Bourouis, M., Jarretou, G., Honegger, B., Paix, A., & Léopold, P. (2008). Drosophila ALS regulates growth and metabolism through functional interaction with insulin-like peptides. *Cell Metabolism*, *7*(4), 333–338. doi:10.1016/j.cmet.2008.02.003

Attwood, T. K. & Findlay, J. B. (1994). Fingerprinting G-protein-coupled receptors. *Protein Eng.* *7*(2), 195–203.

Barberà, M. (2017). *Clock genes and photoperiodism in the aphid Acyrthosiphon pisum* (Doctoral dissertation, Universitat de València).

Barberà, M., Cañas-Cañas, R., & Martínez-Torres, D. (2019). Insulin-like peptides involved in photoperiodism in the aphid *Acyrthosiphon pisum*. *Insect Biochemistry and Molecular Biology*, *112*, 103185. doi:10.1016/j.ibmb.2019.103185

Barberà, M., Collantes-Alegre, J. M., & Martínez-Torres, D. (2017). Characterisation, analysis of expression and localisation of circadian clock genes from the perspective of photoperiodism in the aphid *Acyrthosiphon pisum*. *Insect Biochemistry and Molecular Biology*, *83*, 54–67. doi:https://doi.org/10.1016/j.ibmb.2017.02.006

Barberà, M., Escrivá, L., Collantes-Alegre, J. M., Meca, G., Rosato, E., & Martínez-Torres, D. (2018). Melatonin in the seasonal response of the aphid *Acyrthosiphon pisum*. *Insect Science*. doi:10.1111/1744-7917.12652

Barberà, M. & Martínez-Torres, D. (2017). Identification of the prothoracicotropic hormone (ptth) coding gene and localization of its site of expression in the pea aphid *Acyrthosiphon pisum*. *Insect Molecular Biology*, *26*(5), 654–664. doi:10.1111/imb.12326. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/imb.12326

Barberà, M., Mengual, B., Collantes-Alegre, J. M., Cortés, T., González, A., & Martínez-Torres, D. (2013). Identification, characterization and analysis of expression of genes encoding arylalkylamine N-acetyltransferases in the pea aphid *Acyrthosiphon pisum*. *Insect Molecular Biology*, *22*(6), 623–634. doi:10.1111/imb.12050

Bargiello, T. A., Jackson, F. R., & Young, M. W. (1984). Restoration of circadian behavioural rhythms by gene transfer in drosophila. *Nature*, *312*(5996), 752–754. doi:10.1038/312752a0

Bazalova, O., Kvicalova, M., Valkova, T., Slaby, P., Bartos, P., Netusil, R., ... Vacha, M. (2016). Cryptochrome 2 mediates directional magnetoreception in cockroaches. *Proceedings of the National Academy of Sciences*, *113*(6), 1660–1665. doi:10.1073/pnas.1518622113

Beer, K., Joschinski, J., Arrazola Sastre, A., Krauss, J., & Helfrich-Förster, C. (2017). A damping circadian clock drives weak oscillations in metabolism and locomotor activity of aphids (acyrthosiphon pisum). *Scientific Reports*, *7*(1), 14906. doi:10.1038/s41598-017-15014-3

Bembenek, J., Sehadova, H., Ichihara, N., & Takeda, M. (2005). Day/night fluctuations in melatonin content, arylalkylamine n-acetyltransferase activity and NAT mRNA expression in the CNS, peripheral tissues and hemolymph of the cockroach, *Periplaneta americana*. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, *140*(1), 27–36. doi:10.1016/j.cbpc.2004.03.017

Benito, J., Hoxha, V., Lama, C., Lazareva, A. A., Ferveur, J.-F., Hardin, P. E., & Dauwalder, B. (2010). The circadian output gene takeout

is regulated by pdp1ε. *Proceedings of the National Academy of Sciences*, *107*(6), 2544–2549. doi:10.1073/pnas.0906422107

Bhatt, R. & Chaturvedi, C. M. (1992). The effects of combined serotonin and dopamine precursor application during testicular photosensitivity vs photorefractoriness of the red headed bunting *Emberiza bruniceps*. *Journal of Interdisiplinary Cycle Research*, *23*(2), 120–127. doi:10.1080/09291019209360136

Blackman, R. L. & Eastop, V. F. (2000). *Aphids on the world's crops : An identification and information guide*. Wiley.

Bloch, G., Hazan, E., & Rafaeli, A. (2013). Circadian rhythms and endocrine functions in adult insects. *Journal of Insect Physiology*, *59*(1), 56–69. doi:10.1016/j.jinsphys.2012.10.012

Bodenstein, C., Heiland, I., & Schuster, S. (2012). Temperature compensation and entrainment in circadian rhythms. *Physical Biology*, *9*(3), 036011. doi:10.1088/1478-3975/9/3/036011

Bradshaw, W. E. & Holzapfel, C. M. (2010). Light, time, and the physiology of biotic response to rapid climate change in animals. *Annual Review of Physiology*, *72*(1), 147–166. doi:10.1146/annurev-physiol-021909-135837

Briscoe, A. D. (2001). Functional diversification of lepidopteran opsins following gene duplication. *Molecular Biology and Evolution*, *18*(12), 2270–2279. doi:10.1093/oxfordjournals.molbev.a003773

Briscoe, A. D. (2008). Reconstructing the ancestral butterfly eye: Focus on the opsins. *Journal of Experimental Biology*, *211*(11), 1805–1813. doi:10.1242/jeb.013045

Brisson, J. A., Jaquiery, J., Legeai, F., Le Trionnaire, G., & Tagu, D. (2016). Genomics of phenotypic plasticity in aphids. In *Management of insect pests to agriculture* (pp. 65–96). Springer International Publishing. doi:10.1007/978-3-319-24049-7_3

Bunning, E. (1960). Circadian rhythms and the time measurement in photoperiodism. *Cold Spring Harbor Symposia on Quantitative Biology*, *25*(0), 249–256. doi:10.1101/sqb.1960.025.01.026

Carolan, J. C., Caragea, D., Reardon, K. T., Mutti, N. S., Dittmer, N., Pappan, K., ... Edwards, O. R. (2011). Predicted effector molecules in the salivary secretome of the pea aphid (*Acyrthosiphon pisum*): A dual transcriptomic/proteomic approach. *Journal of Proteome Research*, *10*(4), 1505–1518. doi:10.1021/pr100881q

Chawla, G. & Sokol, N. S. (2012). Hormonal activation of let-7-c microRNAs via EcR is required for adult drosophila melanogaster

morphology and function. *Development*, *139*(10), 1788–1797. doi:10.1242/dev.077743

Chen, W., Liu, Z., Li, T., Zhang, R., Xue, Y., Zhong, Y., ... Zhao, Z. (2014). Regulation of drosophila circadian rhythms by miRNA let-7 is mediated by a regulatory cycle. *Nature Communications*, *5*(1). doi:10.1038/ncomms6549

Ciuk, M. A., Bebas, P., & Cymborowski, B. (2009). Rhythmic oscillations of α-amylase protein and its enzymatic activity levels in Drosophila melanogaster (Diptera: Drosophilidae). *European Journal of Entomology*, *106*(4), 519–528. doi:10.14411/eje.2009.065

Claridge-Chang, A., Wijnen, H., Naef, F., Boothroyd, C., Rajewsky, N., & Young, M. W. (2001). Circadian regulation of gene expression systems in the drosophila head. *Neuron*, *32*(4), 657–671. doi:10.1016/s0896-6273(01)00515-3

Clynen, E., Schoofs, L., & Salzet, M. (2005). A review of the most important classes of serine protease inhibitors in insects and leeches. *Medicinal Chemistry Reviews - Online, Bentham Science*, *3*(2), 197–206.

Colbourne, J. K., Pfrender, M. E., Gilbert, D., Thomas, W. K., Tucker, A., Oakley, T. H., ... Boore, J. L. (2011). The ecoresponsive genome of daphnia pulex. *Science*, *331*(6017), 555–561. doi:10.1126/science.1197761

Cole, C., Barber, J. D., & Barton, G. J. (2008). The jpred 3 secondary structure prediction server. *Nucleic Acids Research*, *36*(Web Server), W197–W201. doi:10.1093/nar/gkn238

Collantes-Alegre, J. M., Mattenberger, F., Barberà, M., & Martínez-Torres, D. (2018). Characterisation, analysis of expression and localisation of the opsin gene repertoire from the perspective of photoperiodism in the aphid *Acyrthosiphon pisum*. *Journal of Insect Physiology*, *104*, 48–59. doi:10.1016/j.jinsphys.2017.11.009

Cook, T. & Desplan, C. (2001). Photoreceptor subtype specification: From flies to humans. *Seminars in Cell & Developmental Biology*, *12*(6), 509–518. doi:10.1006/scdb.2001.0275

Cornelissen, G. (2014). Cosinor-based rhythmometry. *Theoretical Biology and Medical Modelling*, *11*(1). doi:10.1186/1742-4682-11-16

Cortés, T. (2010). *Bases moleculares del control del modo de reproducción en pulgones. identificación de genes regulados por el fotoperiodo y caracterización de los elementos del reloj circadiano*

en Acyrthosiphon pisum (Doctoral dissertation, Universitat de València).

Cortés, T., Ortiz-Rivas, B., & Martínez-Torres, D. (2010). Identification and characterization of circadian clock genes in the pea aphid *Acyrthosiphon pisum*. *Insect Molecular Biology*, *19*, 123–139. doi:10.1111/j.1365-2583.2009.00931.x

Cortés, T., Tagu, D., Simon, J., Moya, A., & Martínez-Torres, D. (2008). Sex versus parthenogenesis: A transcriptomic approach of photoperiod response in the model aphid *Acyrthosiphon pisum* (hemiptera: Aphididae). *Gene*, *408*(1-2), 146–156. doi:10.1016/j.gene.2007.10.030

Cristofoletti, P. T., Ribeiro, A. F., Deraison, C., Rahbé, Y., & Terra, W. R. (2003). Midgut adaptation and digestive enzyme distribution in a phloem feeding insect, the pea aphid *Acyrthosiphon pisum*. *Journal of Insect Physiology*, *49*(1), 11–24. doi:10.1016/s0022-1910(02)00222-6

Cymborowski, B. (1998). Serotonin modulates a photic response in circadian locomotor rhythmicity of adults of the blow fly, *Calliphora vicina*. *Physiological Entomology*, *23*(1), 25–32. doi:10.1046/j.1365-3032.1998.2310025.x

Cyran, S. A., Buchsbaum, A. M., Reddy, K. L., Lin, M.-C., Glossop, N. R., Hardin, P. E., … Blau, J. (2003). Vrille, pdp1, and dClock form a second feedback loop in the drosophila circadian clock. *Cell*, *112*(3), 329–341. doi:10.1016/s0092-8674(03)00074-6

Czarna, A., Berndt, A., Singh, H., Grudziecki, A., Ladurner, A., Timinszky, G., … Wolf, E. (2013). Structures of drosophila cryptochrome and mouse cryptochrome1 provide insight into circadian function. *Cell*, *153*(6), 1394–1405. doi:10.1016/j.cell.2013.05.011

Danks, H. V. (2003). Studying insect photoperiodism and rhythmicity: Components, approaches and lessons. *European Journal of Entomology*, *100*(2), 209–221. doi:10.14411/eje.2003.036. arXiv: eje.2003.036 [DOI: 10.14411]

Dauwalder, B. (2002). The Drosophila takeout gene is regulated by the somatic sexdetermination pathway and affects male courtship behavior. *Genes & Development*, *16*(22), 2879–2892. doi:10.1101/gad.1010302

De Mairan, J. J. d. (1729). Observation botanique. *Histoire de l'Academie Royale des Science*, 35–36.

Denlinger, D., Giebultowicz, J., & Saunders, D. (2001). *Insect timing: Circadian rhythmicity to seasonality* (1st ed.). Elsevier Science.

Dilcher, M. (2003). Use1p is a yeast SNARE protein required for retrograde traffic to the ER. *The EMBO Journal*, *22*(14), 3664–3674. doi:10.1093/emboj/cdg339

Ding, J., Allen, E., Wang, W., Valle, A., Wu, C., Nardine, T., ... Yang, Y. (2006). Gene targeting of GAN in mouse causes a toxic accumulation of microtubule-associated protein 8 and impaired retrograde axonal transport. *Human Molecular Genetics*, *15*(9), 1451–1463. doi:10.1093/hmg/ddl069

Döring, T. F., Kirchner, S. M., Skorupski, P., & Hardie, J. (2011). Spectral sensitivity of the green photoreceptor of winged pea aphids. *Physiological Entomology*, *36*(4), 392–396. doi:10.1111/j.1365-3032.2011.00805.x

Eisenbach, J. & Mittler, T. (1980). An aphid circadian rhythm: Factors affecting the release of sex pheromone by oviparae of the greenbug, *Schizaphis graminum*. *Journal of Insect Physiology*, *26*(8), 511–515. doi:10.1016/0022-1910(80)90125-0

Elekonich, M. M., Schulz, D. J., Bloch, G., & Robinson, G. E. (2001). Juvenile hormone levels in honey bee (*Apis mellifera* l.) foragers: Foraging experience and diurnal variation. *Journal of Insect Physiology*, *47*(10), 1119–1125. doi:10.1016/s0022-1910(01)00090-7

Emery, P., So, W., Kaneko, M., Hall, J. C., & Rosbash, M. (1998). CRY, a drosophila clock and light-regulated cryptochrome, is a major contributor to circadian rhythm resetting and photosensitivity. *Cell*, *95*(5), 669–679. doi:10.1016/s0092-8674(00)81637-2

Emery, P., Stanewsky, R., Helfrich-Förster, C., Emery-Le, M., Hall, J. C., & Rosbash, M. (2000). Drosophila CRY is a deep brain circadian photoreceptor. *Neuron*, *26*(2), 493–504. doi:10.1016/s0896-6273(00)81181-2

Eriksson, B. J., Fredman, D., Steiner, G., & Schmid, A. (2013). Characterisation and localisation of the opsin protein repertoire in the brain and retinas of a spider and an onychophoran. *BMC Evolutionary Biology*, *13*(1), 186. doi:10.1186/1471-2148-13-186

Eskin, A. & Maresh, R. (1982). Serotonin or electrical optic nerve stimulation increases the photosensitivity of the aplysia eye. *Comparative Biochemistry and Physiology Part C: Comparative Pharmacology*, *73*(1), 27–31. doi:10.1016/0306-4492(82)90162-9

Everett, A., Tong, X., Briscoe, A. D., & Monteiro, A. (2012). Phenotypic plasticity in opsin expression in a butterfly compound eye complements sex role reversal. *BMC Evolutionary Biology*, *12*(1), 232. doi:10.1186/1471-2148-12-232

Ewart, G. D., Cannell, D., Cox, G. B., & Howells, A. J. (1994). Mutational analysis of the traffic ATPase (ABC) transporters involved in uptake of eye pigment precursors in Drosophila melanogaster. Implications for structure-function relationships. *J. Biol. Chem.* *269*(14), 10370–10377.

Ewen, A. B. (1962). An improved aldehyde fuchsin staining technique for neurosecretory products in insects. *Transactions of the American Microscopical Society*, *81*(1), 94. doi:10.2307/3223949

Feuda, R., Marlétaz, F., Bentley, M. A., & Holland, P. W. (2016). Conservation, duplication, and divergence of five opsin genes in insect evolution. *Genome Biology and Evolution*, *8*(3), 579–587. doi:10.1093/gbe/evw015

Feyereisen, R. (1999). INSECT p450 ENZYMES. *Annual Review of Entomology*, *44*(1), 507–533. doi:10.1146/annurev.ento.44.1.507

Fine, R., Lehman, W., Head, J., & Blitz, A. (1975). Troponin c in brain. *Nature*, *258*, 260–262.

Finet, C., Slavik, K., Pu, J., Carroll, S. B., & Chung, H. (2019). Birth-and-death evolution of the fatty acyl-CoA reductase (FAR) gene family and diversification of cuticular hydrocarbon synthesis in drosophila. *Genome Biology and Evolution*, *11*(6), 1541–1551. doi:10.1093/gbe/evz094

Follet, B. K. & Sharp, P. J. (1969). Circadian rhythmicity in photoperiodically induced gonadotrophin release and gonadal growth in the quail. *Nature*, *223*(5209), 968–971. doi:10.1038/223968b0

Franke, R., Konig, B., Sakmar, T., Khorana, H., & Hofmann, K. (1990). Rhodopsin mutants that bind but fail to activate transducin. *Science*, *250*(4977), 123–125. doi:10.1126/science.2218504

Fujii-Taira, I., Tanaka, Y., Homma, K. J., & Natori, S. (2000). Hydrolysis and synthesis of substrate proteins for cathepsin l in the brain basement membranes of sarcophaga during metamorphosis. *Journal of Biochemistry*, *128*(3), 539–542. doi:10.1093/oxfordjournals.jbchem.a022784

Futahashi, R., Kawahara-Miki, R., Kinoshita, M., Yoshitake, K., Yajima, S., Arikawa, K., & Fukatsu, T. (2015). Extraordinary diversity of visual opsin genes in dragonflies. *Proceedings of the National*

Academy of Sciences, *112*(11), E1247–E1256. doi:10.1073/pnas.1424670112

Gao, N., Foster, R. G., & Hardie, J. (2000). Two opsin genes from the vetch aphid, *Megoura viciae*. *Insect Mol. Biol. 9*(2), 197–202.

Gao, N., von Schantz, M., Foster, R. G., & Hardie, J. (1999). The putative brain photoperiodic photoreceptors in the vetch aphid, *Megoura viciae*. *Journal of Insect Physiology*, *45*(11), 1011–1019.

Gardner, M., Hubbard, K., Hotta, C., Dodd, A., & Webb, A. R. (2006). How plants tell the time. *Biochemical Journal*, *397*(1), 15–24. doi:10.1042/bj20060484

Garner, W. & Allard, H. (1920). Effect of the relative length of day and night and other factors of the environment on growth and reproduction in plants. *Journal of Agricultural Research*, *18*(11), 553–606.

Gartner, W. & Towner, P. (1995). INVERTEBRATE VISUAL PIGMENTS. *Photochemistry and Photobiology*, *62*(1), 1–16. doi:10.1111/j.1751-1097.1995.tb05231.x

Ge, Z.-Y., Wan, P.-J., Li, G.-Q., Xia, Y.-g., & Han, Z.-J. (2014). Characterization of cysteine protease-like genes in the striped rice stem borer, *Chilo suppressalis*. *Genome*, *57*(2), 79–88. doi:10.1139/gen-2013-0188

Gentile, C., Rivas, G. B. S., Meireles-Filho, A. C. A., Lima, J. B. P., & Peixoto, A. A. (2009). Circadian expression of clock genes in two mosquito disease vectors:cry2is different. *Journal of Biological Rhythms*, *24*(6), 444–451. doi:10.1177/0748730409349169

Goff, L., Trapnell, C., & Kelley, D. (2019). Cummerbund: Analysis, exploration, manipulation, and visualization of cufflinks high-throughput sequencing data. Bioconductor. doi:10.18129/b9.bioc.cummerbund

Goldman, B. (2001). Mammalian photoperiodic system: Formal properties and neuroendocrine mechanisms of photoperiodic time measurement. *Journal of biological rhythms*, *16*, 283–301. doi:10.1177/074873001129001980

Gorbet, D. J. & Steel, C. (2003). A miniature radioimmunoassay for melatonin for use with small samples from invertebrates. *General and Comparative Endocrinology*, *134*(2), 193–197. doi:10.1016/s0016-6480(03)00260-0

Goto, S. G. (2012). Roles of circadian clock genes in insect photoperiodism. *Entomological Science*, *16*(1), 1–16. doi:10.1111/ens.12000

Hansen, K. D., Brenner, S. E., & Dudoit, S. (2010). Biases in illumina transcriptome sequencing caused by random hexamer priming. *Nucleic Acids Research*, *38*(12), e131–e131. doi:10.1093/nar/gkq224

Hardie, J. (1981). Juvenile hormone and photoperiodically controlled polymorphism in aphis fabae: Postnatal effects on presumptive gynoparae. *Journal of Insect Physiology*, *27*(5), 347–355. doi:10.1016/0022-1910(81)90081-0

Hardie, J. (2008). A hormonal basis for the photoperiodic control of polymorphism in aphids. In *Ciba foundation symposium 104 - photoperiodic regulation of insect and molluscan hormones* (Chap. 15, pp. 240–258). John Wiley & Sons, Ltd. doi:10.1002/9780470720851.ch15. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9780470720851.ch15

Hardie, J., Baker, F. C., Jamieson, G. C., A.D. Lees, A. D., & Schooley, D. A. (1985). The identification of an aphid juvenile hormone, and its titre in relation to photoperiod. *Physiological Entomology*, *10*(3), 297–302. doi:10.1111/j.1365-3032.1985.tb00050.x

Hardie, J. & Gao, N. (1997). Melatonin and the pea aphid, *Acyrthosiphon pisum. J. Insect Physiol. 43*(7), 615–620.

Hardin, P. E. (2005). The circadian timekeeping system of drosophila. *Current Biology*, *15*(17), R714–R722. doi:10.1016/j.cub.2005.08.019

Harding, L., Scott, R. H., Kellenberger, C., Hietter, H., Luu, B., Beadle, D. J., & Bermudez, I. (1995). Inhibition of high voltage-activated ca2+ currents from cultured sensory neurones by a novel insect peptide. *Journal of Receptors and Signal Transduction*, *15*(1-4), 355–364. doi:10.3109/10799899509045226

Hegazi, S., Lowden, C., Garcia, J. R., Cheng, A. H., Obrietan, K., Levine, J. D., & Cheng, H.-Y. M. (2019). A symphony of signals: Intercellular and intracellular signaling mechanisms underlying circadian timekeeping in mice and flies. *International Journal of Molecular Sciences*, *20*(9), 2363. doi:10.3390/ijms20092363

Helfrich-Förster, C. (2004). The circadian clock in the brain: A structural and functional comparison between mammals and insects. *Journal of Comparative Physiology A*, *190*(8). doi:10.1007/s00359-004-0527-2

Helfrich-Förster, C. (2019). Light input pathways to the circadian clock of insects with an emphasis on the fruit fly drosophila melanogaster.

Journal of Comparative Physiology A, *206*(2), 259–272. doi:10.1007/s00359-019-01379-5

Henze, M. J., Dannenhauer, K., Kohler, M., Labhart, T., & Gesemann, M. (2012). Opsin evolution and expression in arthropod compound eyes and ocelli: Insights from the cricket *Gryllus bimaculatus*. *BMC Evolutionary Biology*, *12*(1), 163. doi:10.1186/1471-2148-12-163

Hering, L. & Mayer, G. (2014). Analysis of the opsin repertoire in the tardigrade Hypsibius dujardini provides insights into the evolution of opsin genes in Panarthropoda. *Genome Biology and Evolution*, *6*(9), 2380–2391. doi:10.1093/gbe/evu193

Hershko, A. & Ciechanover, A. (1998). THE UBIQUITIN SYSTEM. *Annual Review of Biochemistry*, *67*(1), 425–479. doi:10.1146/annurev.biochem.67.1.425

Hickman, A. B., Klein, D. C., & Dyda, F. (1999). Melatonin Biosynthesis. *Molecular Cell*, *3*(1), 23–32. doi:10.1016/S1097-2765(00)80171-9

Hiragaki, S., Suzuki, T., Mohamed, A. A. M., & Takeda, M. (2015). Structures and functions of insect arylalkylamine n-acetyltransferase (iaaNAT): A key enzyme for physiological and behavioral switch in arthropods. *Frontiers in Physiology*, *6*. doi:10.3389/fphys.2015.00113

Hoffmann, K. H. (2014). *Insect molecular biology and ecology*. CRC Press. Retrieved from https://www.xarg.org/ref/a/1482231883/

Hore, P. J. & Mouritsen, H. (2016). The radical-pair mechanism of magnetoreception. *Annual Review of Biophysics*, *45*(1), 299–344. doi:10.1146/annurev-biophys-032116-094545

Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, *4*(1), 44–57. doi:10.1038/nprot.2008.211

Hwa, J., Reeves, P. J., Klein-Seetharaman, J., Davidson, F., & Khorana, H. G. (1999). Structure and function in rhodopsin: Further elucidation of the role of the intradiscal cysteines, cys-110, -185, and -187, in rhodopsin folding and function. *Proceedings of the National Academy of Sciences*, *96*(5), 1932–1935. doi:10.1073/pnas.96.5.1932

Ikegami, K. & Yoshimura, T. (2012). Circadian clocks and the measurement of daylength in seasonal reproduction. *Molecular and Cellular Endocrinology*, *349*(1), 76–81. doi:10.1016/j.mce.2011.06.040

Isabel, G., Gourdoux, L., & Moreau, R. (2001). Changes of biogenic amine levels in haemolymph during diapausing and non-diapausing status

in pieris brassicae l. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, *128*(1), 117–127. doi:10.1016/s1095-6433(00)00284-1

Ishikawa, A., Ogawa, K., Gotoh, H., Walsh, T. K., Tagu, D., Brisson, J. A., ... Miura, T. (2011). Juvenile hormone titre and related gene expression during the change of reproductive modes in the pea aphid. *Insect Molecular Biology*, *21*(1), 49–60. doi:10.1111/j.1365-2583.2011.01111.x

Jarwar, A. R., Hao, K., Bitume, E. V., Ullah, H., Cui, D., Nong, X., ... Zhang, Z. (2019). Comparative transcriptomic analysis reveals molecular profiles of central nervous system in maternal diapause induction of locusta migratoria. *G3: GENES, GENOMES, GENETICS*, *9*(10), 3287–3296. doi:10.1534/g3.119.400475

Jaubert-Possamai, S., Le Trionnaire, G., Bonhomme, J., Christophides, G. K., Rispe, C., & Tagu, D. (2007). Gene knockdown by RNAi in the pea aphid acyrthosiphon pisum. *BMC Biotechnology*, *7*(1), 63. doi:10.1186/1472-6750-7-63

Jenkins, A. M. & Muskavitch, M. A. T. (2015). Crepuscular behavioral variation and profiling of opsin genes in *Anopheles gambiae* and *Anopheles stephensi* (diptera: Culicidae). *Journal of Medical Entomology*, *52*(3), 296–307. doi:10.1093/jme/tjv024

Jeon, H., Han, K. S., & Boo, K. S. (2003). Sex pheromone of aphis spiraecola (homoptera: Aphididae): Composition and circadian rhythm in release. *Journal of Asia-Pacific Entomology*, *6*(2), 159–165. doi:10.1016/s1226-8615(08)60181-8

Ji, R., Wang, Y., Cheng, Y., Zhang, M., Zhang, H.-B., Zhu, L., ... Zhu-Salzman, K. (2016). Transcriptome analysis of green peach aphid (myzus persicae): Insight into developmental regulation and inter-species divergence. *Frontiers in Plant Science*, *7*. doi:10.3389/fpls.2016.01562

Johnston, J. R., Chase, P. B., & Pinto, J. R. (2017). Troponin through the looking-glass: Emerging roles beyond regulation of striated muscle contraction. *Oncotarget*, *9*(1). doi:10.18632/oncotarget.22879

Kanost, M. R. & Clem, R. J. (2017). Insect proteases. In *Reference module in life sciences*. Elsevier. doi:10.1016/b978-0-12-809633-8.04046-2

Katanosaka, K., Tokunaga, F., Kawamura, S., & Ozaki, K. (1998). N -linked glycosylation of drosophila rhodopsin occurs exclusively in the amino-terminal domain and functions in rhodopsin maturation.

FEBS Letters, *424*(3), 149–154. doi:10.1016/s0014-5793(98)00160-4

Kijimoto, T., Moczek, A. P., & Andrews, J. (2012). Diversification of doublesex function underlies morph-, sex-, and species-specific development of beetle horns. *Proceedings of the National Academy of Sciences*, *109*(50), 20526–20531. doi:10.1073/pnas.1118589109. eprint: https://www.pnas.org/content/109/50/20526.full.pdf

Kim, D., Langmead, B., & Salzberg, S. L. (2015). HISAT: A fast spliced aligner with low memory requirements. *Nature Methods*, *12*(4), 357–360. doi:10.1038/nmeth.3317

King, D. P., Vitaterna, M. H., Chang, A.-M., Dove, W. F., Pinto, L. H., Turek, F. W., & Takahashi, J. S. (1997a). The mouse *Clock* mutation behaves as an antimorph and maps within the w19h deletion, distal of kit. *Genetics*, *146*(3), 1049–1060.

King, D. P., Zhao, Y., Sangoram, A. M., Wilsbacher, L. D., Tanaka, M., Antoch, M. P., … Takahashi, J. S. (1997b). Positional cloning of the mouse circadian *Clock* gene. *Cell*, *89*(4), 641–653.

Kirchner, S., Döring, T., & Saucke, H. (2005). Evidence for trichromacy in the green peach aphid, myzus persicae (sulz.) (hemiptera: Aphididae). *Journal of Insect Physiology*, *51*(11), 1255–1260. doi:10.1016/j.jinsphys.2005.07.002

Kistenpfennig, C. R. (2012). *Rhodopsin 7 and cryptochrome – circadian photoreception in Drosophila* (Doctoral dissertation, Julius-Maximilians-Universität Wurzburg).

Kjer, K. M., Simon, C., Yavorskaya, M., & Beutel, R. G. (2016). Progress, pitfalls and parallel universes: A history of insect phylogenetics. *Journal of The Royal Society Interface*, *13*(121), 20160363. doi:10.1098/rsif.2016.0363

Klebs, G. (1910). Alterations in the development and forms of plants as a result of environment. *Proc. Roy. Soc. Lond. B*, *82*, 547–558.

Klebs, G. (1913). Über das verhältnis der aussenwelt zur entwicklung der pflanze [on the relationship of the outside world to the development of the plant]. *Sber. Akad. Wiss. Heidelberg*, *5*, 1–47.

Klein, D. C. (2006). Arylalkylamine-N-acetyltransferase: "the Timezyme". *Journal of Biological Chemistry*, *282*(7), 4233–4237. doi:10.1074/jbc.r600036200

Kobe, B. & Deisenhofer, J. (1994). The leucine-rich repeat: A versatile binding motif. *Trends in Biochemical Sciences*, *19*(10), 415–421. doi:10.1016/0968-0004(94)90090-6

Koh, K. (2006). JETLAG resets the drosophila circadian clock by promoting light-induced degradation of TIMELESS. *Science*, *312*(5781), 1809–1812. doi:10.1126/science.1124951

Kojima, S. & Green, C. B. (2014). Circadian genomics reveal a role for post-transcriptional regulation in mammals. *Biochemistry*, *54*(2), 124–133. doi:10.1021/bi500707c

Kollmann, M., Minoli, S., Bonhomme, J., Homberg, U., Schachtner, J., Tagu, D., & Anton, S. (2010). Revisiting the anatomy of the central nervous system of a hemimetabolous model insect species: The pea aphid acyrthosiphon pisum. *Cell and Tissue Research*, *343*(2), 343–355. doi:10.1007/s00441-010-1099-9

Koning, R. E. (1994). Photoperiodism. plant physiology information website. Retrieved October 14, 2019, from http://plantphys.info/plant_physiology/photoperiodism.shtml

Konopka, R. J. & Benzer, S. (1971). Clock mutants of drosophila melanogaster. *Proceedings of the National Academy of Sciences*, *68*(9), 2112–2116. doi:10.1073/pnas.68.9.2112

Koštál, V. (2011). Insect photoperiodic calendar and circadian clock: Independence, cooperation, or unity? *Journal of Insect Physiology*, *57*(5), 538–556. doi:10.1016/j.jinsphys.2010.10.006

Krueger, F. (2015). Trim galore. a wrapper tool around cutadapt and fastqc to consistently apply quality and adapter trimming to fastq files. Retrieved from https://github.com/FelixKrueger/TrimGalore

Kuhlman, S. J., Craig, L. M., & Duffy, J. F. (2017). Introduction to chronobiology. *Cold Spring Harbor Perspectives in Biology*, *10*(9), a033613. doi:10.1101/cshperspect.a033613

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA x: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, *35*(6), 1547–1549. doi:10.1093/molbev/msy096

Kutsukake, M., Shibao, H., Nikoh, N., Morioka, M., Tamura, T., Hoshino, T., ... Fukatsu, T. (2004). Venomous protease of aphid soldier for colony defense. *Proceedings of the National Academy of Sciences*, *101*(31), 11338–11343. doi:10.1073/pnas.0402462101

Landolt, P. J., Reed, H. C., & Heath, R. R. (1999). An alarm pheromone from heads of worker *Vespula squamosa* (hymenoptera: Vespidae). *The Florida Entomologist*, *82*(2), 356. doi:10.2307/3496590

Le Trionnaire, G., Francis, F., Jaubert-Possamai, S., Bonhomme, J., Pauw, E. D., Gauthier, J.-P., ... Tagu, D. (2009). Transcriptomic

and proteomic analyses of seasonal photoperiodism in the pea aphid. *BMC Genomics*, *10*(1), 456. doi:10.1186/1471-2164-10-456

Le Trionnaire, G., Jaubert, S., Sabater-Muñoz, B., Benedetto, A., Bonhomme, J., Prunier-Leterme, N., ... Tagu, D. (2007). Seasonal photoperiodism regulates the expression of cuticular and signalling protein genes in the pea aphid. *Insect Biochemistry and Molecular Biology*, *37*(10), 1094–1102. doi:10.1016/j.ibmb.2007.06.008

Le Trionnaire, G., Jaubert-Possamai, S., Bonhomme, J., Gauthier, J.-P., Guernec, G., Cam, A. L., ... Tagu, D. (2012). Transcriptomic profiling of the reproductive mode switch in the pea aphid in response to natural autumnal photoperiod. *Journal of Insect Physiology*, *58*(12), 1517–1524. doi:10.1016/j.jinsphys.2012.07.009

LeBlanc, G. A. & Medlock, E. K. (2015). Males on demand: the environmental-neuro-endocrine control of male sex determination in daphnids. *FEBS Journal*, *282*(21), 4080–4093. doi:10.1111/febs.13393

Lees, A. (1964). The location of the photoperiodic receptors in the aphid *Megoura viciae* Buckton. *Journal of Experimental Biology*, *41*, 119–133.

Lees, A. (1973). Photoperiodic time measurement in the aphid *Megoura viciae*. *Journal of Insect Physiology*, *19*(12), 2279–2316. doi:10.1016/0022-1910(73)90237-0

Lees, A. (1981). Action spectra for the photoperiodic control of polymorphism in the aphid *Megoura viciae*. *Journal of Insect Physiology*, *27*(11), 761–771. doi:10.1016/0022-1910(81)90066-4

Lewis, R. & Saunders, D. (1987). A damped circadian oscillator model of an insect photoperiodic clock. i. description of the model based on a feedback control system. *Journal of Theoretical Biology*, *128*(1), 47–59. doi:https://doi.org/10.1016/S0022-5193(87)80030-9

Li, C., Yun, X., Hu, X., Zhang, Y., Sang, M., Liu, X., ... Li, B. (2013). Identification of g protein-coupled receptors in the pea aphid, acyrthosiphon pisum. *Genomics*, *102*(4), 345–354. doi:10.1016/j.ygeno.2013.06.003

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... and, R. D. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. doi:10.1093/bioinformatics/btp352

Lienard, M. A., Hagstrom, A. K., Lassance, J.-M., & Lofstedt, C. (2010). Evolution of multicomponent pheromone signals in small ermine moths involves a single fatty-acyl reductase gene. *Proceedings of*

the National Academy of Sciences, *107*(24), 10955–10960. doi:10.1073/pnas.1000823107

Lin, F.-J., Song, W., Meyer-Bernstein, E., Naidoo, N., & Sehgal, A. (2001). Photic signaling by cryptochrome in the DrosophilaCircadian system. *Molecular and Cellular Biology*, *21*(21), 7287–7294. doi:10.1128/mcb.21.21.7287-7294.2001

Livak, K. J. & Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods San Diego Calif*, *25*(4), 402–408. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11846609

Lomberk, G., Wallrath, L., & Urrutia, R. (2006). The heterochromatin protein 1 family. *Genome Biology*, *7*(7), 228. doi:10.1186/gb-2006-7-7-228

Lyckman, A. W., Horng, S., Leamey, C. A., Tropea, D., Watakabe, A., Wart, A. V., ... Sur, M. (2008). Gene expression patterns in visual cortex during the critical period: Synaptic stabilization and reversal by visual deprivation. *Proceedings of the National Academy of Sciences*, *105*(27), 9409–9414. doi:10.1073/pnas.0710172105

MacKay, P. A. (1987). Production of sexual and asexual morphs and changes in reproductive sequence associated with photoperiod in the pea aphid, *Acyrthosiphon pisum* (harris). *Canadian Journal of Zoology*, *65*(11), 2602–2606. doi:10.1139/z87-394

Mackay, P. A. (1989). Clonal variation in sexual morph production in *Acyrthosiphon pisum* (homoptera: Aphididae). *Environmental Entomology*, *18*(4), 558–562. doi:10.1093/ee/18.4.558

Majercak, J., Sidote, D., Hardin, P. E., & Edery, I. (1999). How a circadian clock adapts to seasonal decreases in temperature and day length. *Neuron*, *24*(1), 219–230. doi:10.1016/s0896-6273(00)80834-x

Marcovitch, S. (1923). Plant lice and light exposure. *Science*, *58*(1513), 537–538. doi:10.1126/science.58.1513.537-a

Marcovitch, S. (1924). The migration of the aphididae and the appearance of the sexual forms as affected by the relative length of daily light exposure. *J. agric. res. (Wash. D.C.). Journal of agricultural research. 27*(7), 513–522.

Matthews, J. M. & Sunde, M. (2002). Zinc fingers–folds for many occasions. *IUBMB Life (International Union of Biochemistry and Molecular Biology: Life)*, *54*(6), 351–355. doi:10.1080/15216540216035

Mei, Q. & Dvornyk, V. (2015). Evolutionary history of the photolyase/cryptochrome superfamily in eukaryotes. *PLOS ONE*, *10*(9), e0135940. doi:10.1371/journal.pone.0135940

Messina, G., Celauro, E., Atterrato, M. T., Giordano, E., Iwashita, S., & Dimitri, P. (2014a). The bucentaur (BCNT) protein family: A long-neglected class of essential proteins required for chromatin/chromosome organization and function. *Chromosoma*, *124*(2), 153–162. doi:10.1007/s00412-014-0503-8

Messina, G., Damia, E., Fanti, L., Atterrato, M. T., Celauro, E., Mariotti, F. R., ... Dimitri, P. (2014b). Yeti, an essential drosophila melanogaster gene, encodes a protein required for chromatin organization. *Journal of Cell Science*, *127*(11), 2577–2588. doi:10.1242/jcs.150243

Meuti, M. E., Stone, M., Ikeno, T., & Denlinger, D. (2015). Functional circadian clock genes are essential for the overwintering diapause of the northern house mosquito, culex pipiens. *Journal of Experimental Biology*, *218*(3), 412–422. doi:10.1242/jeb.113233

Michael, A. K., Fribourgh, J. L., Gelder, R. N. V., & Partch, C. L. (2017). Animal cryptochromes: Divergent roles in light perception, circadian timekeeping and beyond. *Photochemistry and Photobiology*, *93*(1), 128–140. doi:10.1111/php.12677

Mirzadegan, T., Benkö, G., Filipek, S., & Palczewski, K. (2003). Sequence analyses of G-protein-coupled receptors: similarities to rhodopsin. *Biochemistry*, *42*(10), 2759–2767.

Miyazaki, M. (1987). Morphology and systematics. In *World crop pests - aphids their biology, natural enemies and control* (Chap. 1, Vol. 2A, pp. 1–23). Elsevier Science Publishers B.V.

Miyazaki, Y., Nisimura, T., & Numata, H. (2005). A phase response curve for circannual rhythm in the varied carpet beetle anthrenus verbasci. *Journal of Comparative Physiology A*, *191*(10), 883–887. doi:10.1007/s00359-005-0012-6

Mohamed, A. A. M., Wang, Q., Bembenek, J., Ichihara, N., Hiragaki, S., Suzuki, T., & Takeda, M. (2014). N-acetyltransferase (nat) is a critical conjunct of photoperiodism between the circadian system and endocrine axis in *Antheraea pernyi*. *PLoS ONE*, *9*(3), e92680. doi:10.1371/journal.pone.0092680

Moran, N. A. & Jarvik, T. (2010). Lateral transfer of genes from fungi underlies carotenoid production in aphids. *Science*, *328*(5978), 624–627. doi:10.1126/science.1187113

Murray, A. R., Fliesler, S. J., & Al-Ubaidi, M. R. (2009). Rhodopsin: The functional significance of asn-linked glycosylation and other post-translational modifications. *Ophthalmic Genetics*, *30*(3), 109–120. doi:10.1080/13816810902962405

Nakabachi, A., Shigenobu, S., Sakazume, N., Shiraki, T., Hayashizaki, Y., Carninci, P., ... Fukatsu, T. (2005). Transcriptome analysis of the aphid bacteriocyte, the symbiotic host cell that harbors an endocellular mutualistic bacterium, buchnera. *Proceedings of the National Academy of Sciences*, *102*(15), 5477–5482. doi:10.1073/pnas.0409034102

Narayandas, G. K. & Alyokhin, A. V. (2006). Diurnal patterns in host finding by potato aphids, macrosiphum euphorbiae (homoptera: Aphididae). *Journal of Insect Behavior*, *19*(3), 347. doi:10.1007/s10905-006-9029-0

Nei, M. & Kumar, S. (2000). *Molecular evolution and phylogenetics* (1st ed.). Oxford University Press, USA.

Nelson, R. E., Fessler, L. I., Takagi, Y., Blumberg, B., Keene, D. R., Olson, P. F., ... Fessler, J. H. (1994). Peroxidasin: a novel enzyme-matrix protein of Drosophila development. *The EMBO journal*, *13*(15), 3438–47. doi:10.1002/j.1460-2075.1994.tb06649.x

Nelson, R., Denlinger, D., & Somers, D. (2010). *Photoperiodism: The biological calendar*. Oxford University Press. Retrieved from https://books.google.es/books?id=RCpnkiegrTYC

Neville, A. (1983). Daily cuticular growth layers and the teneral stage in adult insects: A review. *Journal of Insect Physiology*, *29*(3), 211–219. doi:10.1016/0022-1910(83)90087-2

Ni, J. D., Baik, L. S., Holmes, T. C., & Montell, C. (2017). A rhodopsin in the brain functions in circadian photoentrainment in drosophila. *Nature*, *545*(7654), 340–344. doi:10.1038/nature22325

Nicholas, K. & Nicholas, H. B. (1997). Genedoc: A tool for editing and annotating multiple sequence alignments.

Noguchi, H. & Hayakawa, Y. (1997). Role of dopamine at the onset of pupal diapause in the cabbage armyworm *Mamestra brassicae*. *FEBS Lett. 413*(1), 157–161.

Nunes, M. V. & Hardie, J. (1993). Circadian rhythmicity is involved in photoperiodic time measurement in the aphid *Megoura viciae*. *Experientia*, *49*(8), 711–713. doi:10.1007/bf01923957

Oba, Y. & Kainuma, T. (2009). Diel changes in the expression of long wavelength-sensitive and ultraviolet-sensitive opsin genes in the

japanese firefly, *Luciola cruciata*. *Gene*, *436*(1-2), 66–70. doi:10.1016/j.gene.2009.02.001

Oerke, E.-C. & Dehne, H.-W. (1997). Global crop production and the efficacy of crop protection - current situation and future trends. *European Journal of Plant Pathology*, *103*(3), 203–215. doi:10.1023/A:1008602111248

Ogawa, K. & Miura, T. (2014). Aphid polyphenisms: Trans-generational developmental regulation through viviparity. *Frontiers in Physiology*, *5 JAN*(January), 1–11. doi:10.3389/fphys.2014.00001

Okonechnikov, K., Golosova, O., & Fursov, M. (2012). Unipro UGENE: A unified bioinformatics toolkit. *Bioinformatics*, *28*(8), 1166–1167. doi:10.1093/bioinformatics/bts091

O'Tousa, J. E. (1992). Requirement of n-linked glycosylation site in *Drosophila* rhodopsin. *Visual Neuroscience*, *8*(5), 385–390. doi:10.1017/S0952523800004910

O'Tousa, J. E., Baehr, W., Martin, R. L., Hirsh, J., Pak, W. L., & Applebury, M. L. (1985). The *Drosophila* ninaE gene encodes an opsin. *Cell*, *40*(4), 839–850. doi:10.1016/0092-8674(85)90343-5

Ouzzine, M., Gulberti, S., Ramalanjaona, N., Magdalou, J., & Fournel-Gigleux, S. (2014). The UDP-glucuronosyltransferases of the blood-brain barrier: their role in drug metabolism and detoxication. *Frontiers in Cellular Neuroscience*, *8*. doi:10.3389/fncel.2014.00349

Ozturk, N., VanVickle-Chavez, S. J., Akileswaran, L., Gelder, R. N. V., & Sancar, A. (2013). Ramshackle (brwd3) promotes light-induced ubiquitylation of *Drosophila* cryptochrome by DDB1-CUL4-ROC1 e3 ligase complex. *Proceedings of the National Academy of Sciences*, *110*(13), 4980–4985. doi:10.1073/pnas.1303234110

Page, T. L. (1987). Serotonin phase-shifts the circadian rhythm of locomotor activity in the cockroach. *Journal of Biological Rhythms*, *2*(1), 23–34. doi:10.1177/074873048700200103

Peschel, N., Chen, K. F., Szabo, G., & Stanewsky, R. (2009). Light-dependent interactions between the drosophila circadian clock factors cryptochrome, jetlag, and timeless. *Current Biology*, *19*(3), 241–247. doi:10.1016/j.cub.2008.12.042

Pesquita, C., Faria, D., Falcão, A. O., Lord, P., & Couto, F. M. (2009). Semantic similarity in biomedical ontologies. *PLoS Computational Biology*, *5*(7), e1000443. doi:10.1371/journal.pcbi.1000443

Piskounova, E., Polytarchou, C., Thornton, J., LaPierre, R., Pothoulakis, C., Hagan, J., ... Gregory, R. (2011). Lin28a and lin28b inhibit

let-7 MicroRNA biogenesis by distinct mechanisms. *Cell*, *147*(5), 1066–1079. doi:10.1016/j.cell.2011.10.039

Pittendrigh, C. S. & Minis, D. H. (1964). The entrainment of circadian oscillations by light and their role as photoperiodic clocks. *The American Naturalist*, *98*(902), 261–294. Retrieved from http://www.jstor.org/stable/2459454

Pittendrigh, C. (1966). The circadian oscillation in drosophila pseudoobscura pupae: A model for the photoperiodic clock. *Zeitschrift fur Pflanzenphysiologie*, (54), 275–307.

Plachetzki, D. C., Degnan, B. M., & Oakley, T. H. (2007). The origins of novel protein interactions during animal opsin evolution. *PLoS ONE*, *2*(10), e1054. doi:10.1371/journal.pone.0001054

Porter, M. L., Blasic, J. R., Bok, M. J., Cameron, E. G., Pringle, T., Cronin, T. W., & Robinson, P. R. (2011). Shedding new light on opsin evolution. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1726), 3–14. doi:10.1098/rspb.2011.1819

Posnien, N., Hopfen, C., Hilbrant, M., Ramos-Womack, M., Murat, S., Schönauer, A., ... McGregor, A. P. (2012). Evolution of eye morphology and rhodopsin expression in the drosophila melanogaster species subgroup. *PLoS ONE*, *7*(5), e37346. doi:10.1371/journal.pone.0037346

Prado, F., Jimeno-González, S., & Reyes, J. C. (2017). Histone availability as a strategy to control gene expression. *RNA Biology*, *14*(3), 281–286. doi:10.1080/15476286.2016.1189071

Price, D., Karley, A., Ashford, D., Isaacs, H., Pownall, M., Wilkinson, H., ... Douglas, A. (2007). Molecular characterisation of a candidate gut sucrase in the pea aphid, *Acyrthosiphon pisum*. *Insect Biochemistry and Molecular Biology*, *37*(4), 307–317. doi:10.1016/j.ibmb.2006.12.005

Provencio, I., Jiang, G., Grip, W. J. D., Hayes, W. P., & Rollag, M. D. (1998). Melanopsin: An opsin in melanophores, brain, and eye. *Proceedings of the National Academy of Sciences*, *95*(1), 340–345. doi:10.1073/pnas.95.1.340

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.

R Core Team. (2014). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. Retrieved from http://www.R-project.org/

Rabatel, A., Febvay, G., Gaget, K., Duport, G., Baa-Puyoulet, P., Sapountzis, P., ... Colella, S. (2013). Tyrosine pathway regulation is host-mediated in the pea aphid symbiosis during late embryonic and early larval development. *BMC Genomics*, *14*(1), 235. doi:10.1186/1471-2164-14-235

Rattray, A. M. M. & Müller, B. (2012). The control of histone gene expression. *Biochemical Society Transactions*, *40*(4), 880–885. doi:10.1042/BST20120065

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. doi:10.1016/0092-8674(84)90265-4

Renn, S. C., Park, J. H., Rosbash, M., Hall, J. C., & Taghert, P. H. (1999). A pdf neuropeptide gene mutation and ablation of PDF neurons each cause severe abnormalities of behavioral circadian rhythms in drosophila. *Cell*, *99*(7), 791–802. doi:10.1016/s0092-8674(00)81676-1

Reznik, S. Y., Dolgovskaya, M. Y., & Ovchinnikov, A. N. (2015). Effect of photoperiod on adult size and weight in *Harmonia axyridis* (coleoptera: Coccinellidae). *European Journal of Entomology*, *112*(4), 642–647. doi:10.14411/eje.2015.081

Rivas, G. B. S., da R. Bauzer, L. G. S., & Meireles-Filho, A. C. A. (2016). "The environment is everything that isn't me": Molecular mechanisms and evolutionary dynamics of insect clocks in variable surroundings. *Frontiers in Physiology*, *6*. doi:10.3389/fphys.2015.00400

Roubaud, E. (1918). Rythmes physiologiques et vol spontané chez l'anopheles maculipennis. *CR Hebdomadaires des Séances de l'Academie des Science Paris*, *167*, 967–969.

Rubin, E. B., Shemesh, Y., Cohen, M., Elgavish, S., Robertson, H. M., & Bloch, G. (2006). Molecular and phylogenetic analyses reveal mammalian-like clockwork in the honey bee (*Apis mellifera*) and shed new light on the molecular evolution of the circadian clock. *Genome Research*, *16*(11), 1352–1365. doi:10.1101/gr.5094806

Rund, S. S. C., Hou, T. Y., Ward, S. M., Collins, F. H., & Duffield, G. E. (2011). Genome-wide profiling of diel and circadian gene expression in the malaria vectorAnopheles gambiae. *Proceedings*

of the National Academy of Sciences, *108*(32), E421–E430. doi:10.1073/pnas.1100584108

Rund, S. S. C., Yoo, B., Alam, C., Green, T., Stephens, M. T., Zeng, E., ... Pfrender, M. E. (2016). Genome-wide profiling of 24 hr diel rhythmicity in the water flea, daphnia pulex: Network analysis reveals rhythmic gene expression and enhances functional gene annotation. *BMC Genomics*, *17*(1). doi:10.1186/s12864-016-2998-2

Sabater-Muñoz, B., Legeai, F., Rispe, C., Bonhomme, J., Dearden, P., Dossat, C., ... Tagu, D. (2006). *Genome Biology*, *7*(3), R21. doi:10.1186/gb-2006-7-3-r21

Saikhedkar, N., Summanwar, A., Joshi, R., & Giri, A. (2015). Cathepsins of lepidopteran insects: Aspects and prospects. *Insect Biochemistry and Molecular Biology*, *64*, 51–59. doi:10.1016/j.ibmb.2015.07.005

Saint-Charles, A., Michard-Vanhée, C., Alejevski, F., Chélot, E., Boivin, A., & Rouyer, F. (2016). Four of the six drosophila rhodopsin-expressing photoreceptors can mediate circadian entrainment in low light. *Journal of Comparative Neurology*, *524*(14), 2828–2844. doi:10.1002/cne.23994

Salazar, A., Fürstenau, B., Quero, C., Pérez-Hidalgo, N., Carazo, P., Font, E., & Martínez-Torres, D. (2015). Aggressive mimicry coexists with mutualism in an aphid. *Proceedings of the National Academy of Sciences*, *112*(4), 1101–1106. doi:10.1073/pnas.1414061112

Salcedo, E., Zheng, L., Phistry, M., Bagg, E. E., & Britt, S. G. (2003). Molecular basis for ultraviolet vision in invertebrates. *The Journal of Neuroscience*, *23*(34), 10873–10878. doi:10.1523/jneurosci.23-34-10873.2003

Sanes, J. R. & Zipursky, S. L. (2010). Design principles of insect and vertebrate visual systems. *Neuron*, *66*(1), 15–36. doi:10.1016/j.neuron.2010.01.018

Sarfare, S., Ahmad, S. T., Joyce, M. V., Boggess, B., & O'Tousa, J. E. (2005). The drosophila ninaG oxidoreductase acts in visual pigment chromophore production. *Journal of Biological Chemistry*, *280*(12), 11895–11901. doi:10.1074/jbc.m412236200

Sarov-Blat, L., So, W. V., Liu, L., & Rosbash, M. (2000). The drosophila takeout gene is a novel molecular link between circadian rhythms and feeding behavior. *Cell*, *101*(6), 647–656. doi:10.1016/s0092-8674(00)80876-4

Sasagawa, H., Narita, R., Kitagawa, Y., & Kadowaki, T. (2003). The expression of genes encoding visual components is regulated by

a circadian clock, light environment and age in the honeybee (*Apis mellifera*). *European Journal of Neuroscience*, *17*(5), 963–970. doi:10.1046/j.1460-9568.2003.02528.x

Saunders, D. (2002). *Insect Clocks, Third Edition*. Elsevier Science.

Saunders, D. (2009). Circadian rhythms and the evolution of photoperiodic timing in insects. *Physiological Entomology*, *34*(4), 301–308. doi:10.1111/j.1365-3032.2009.00699.x

Saunders, D. (2012). Insect photoperiodism: Seeing the light. *Physiological Entomology*, *37*(3), 207–218. doi:10.1111/j.1365-3032.2012.00837.x

Schuler, M. A. & Sligar, S. G. (2007). Diversities and similarities in p450 systems: An introduction. In *The ubiquitous roles of cytochrome p450 proteins* (pp. 1–26). John Wiley & Sons, Ltd. doi:10.1002/9780470028155.ch1

Schumann, I., Hering, L., & Mayer, G. (2016). Immunolocalization of arthropsin in the onychophoran euperipatoides rowelli (peripatopsidae). *Frontiers in Neuroanatomy*, *10*. doi:10.3389/fnana.2016.00080

Sehgal, A., Price, J., Man, B., & Young, M. (1994). Loss of circadian behavioral rhythms and per RNA oscillations in the drosophila mutant timeless. *Science*, *263*(5153), 1603–1606. doi:10.1126/science.8128246

Senthilan, P. R., Grebler, R., Reinhard, N., Rieger, D., & Helfrich-Förster, C. (2019). Role of rhodopsins as circadian photoreceptors in the drosophila melanogaster. *Biology*, *8*(1), 6. doi:10.3390/biology8010006

Senthilan, P. R. & Helfrich-Förster, C. (2016). Rhodopsin 7–the unusual rhodopsin inDrosophila. *PeerJ*, *4*, e2427. doi:10.7717/peerj.2427

Shafer, O. T., Levine, J. D., Truman, J. W., & Hall, J. C. (2004). Flies by night: Effects of changing day length on drosophila's circadian clock. *Current Biology*, *14*(5), 424–432. doi:10.1016/j.cub.2004.02.038

Shafer, O. T. & Yao, Z. (2014). Pigment-dispersing factor signaling and circadian rhythms in insect locomotor activity. *Current Opinion in Insect Science*, *1*, 73–80. doi:10.1016/j.cois.2014.05.002

Shen, W. L., Kwon, Y., Adegbola, A. A., Luo, J., Chess, A., & Montell, C. (2011). Function of rhodopsin in temperature discrimination in drosophila. *Science*, *331*(6022), 1333–1336. doi:10.1126/science.1198904

Shichida, Y., Yamashita, T., Imai, H., & Kishida, T. (2013). *Evolution and senses*. Springer Japan. doi:10.1007/978-4-431-54222-3

Sim, C. & Denlinger, D. (2009). A shut-down in expression of an insulin-like peptide, ILP-1, halts ovarian maturation during the overwintering diapause of the mosquito *Culex pipiens*. *Insect Molecular Biology*, *18*(3), 325–332. doi:10.1111/j.1365-2583.2009.00872.x

Sim, C. & Denlinger, D. (2013). Insulin signaling and the regulation of insect diapause. *Frontiers in Physiology*, *4*. doi:10.3389/fphys.2013.00189

Sim, C., Kang, D. S., Kim, S., Bai, X., & Denlinger, D. (2015). Identification of foxo targets that generate diverse features of the diapause phenotype in the mosquito *Culex pipiens*. *Proceedings of the National Academy of Sciences*, 201502751. doi:10.1073/pnas.1502751112

Smith, M. A. H. & MacKay, P. A. (1990). Latitudinal variation in the photoperiodic responses of populations of pea aphid (homoptera: Aphididae). *Environmental Entomology*, *19*(3), 618–624. doi:10.1093/ee/19.3.618

So, W. V., Sarov-Blat, L., Kotarski, C. K., McDonald, M. J., Allada, R., & Rosbash, M. (2000). Takeout, a novel drosophila gene under circadian clock transcriptional regulation. *Molecular and cellular biology*, *20*(18), 6935–6944. doi:10.1128/mcb.20.18.6935-6944.2000

Soudi, M., Zamocky, M., Jakopitsch, C., Furtmüller, P. G., & Obinger, C. (2012). Molecular Evolution, Structure, and Function of Peroxidasins. *Chemistry & Biodiversity*, *9*(9), 1776–1793. doi:10.1002/cbdv.201100438

Staden, R., Beal, K. F., & Bonfield, J. K. (1998). The staden package, 1998. In *Bioinformatics methods and protocols* (pp. 115–130). Humana Press. doi:10.1385/1-59259-192-2:115

Steel, C. (1976). NEUROSECRETORY CONTROL OF POLYMORPHISM IN APHIDS. In *Phase and caste determination in insects* (pp. 117–130). Elsevier. doi:10.1016/b978-0-08-021256-2.50015-7

Steel, C. (1977). The neurosecretory system in the aphid *Megoura viciae*, with reference to unusual features associated with long distance transport of neurosecretion. *General and Comparative Endocrinology*, *31*(3), 307–322. doi:10.1016/0016-6480(77)90095-8

Steel, C. (1978). Some functions of identified neurosecretory cells in the brain of the aphid, *Megoura viciae*. *General and Comparative Endocrinology*, *34*(2), 219–228. doi:10.1016/0016-6480(78)90213-7

Steel, C. & Lees, A. (1977). The role of neurosecretion in the photoperiodic control of polymorphism in the aphid *Megoura viciae*. *The Journal*

of experimental biology, *67*, 117–35. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/894176

Steel, C. & Vafopoulou, X. (2002). Physiology of circadian systems. In C. Steel, X. Vafopoulou, & R. Lewis (Eds.), *Insect clocks* (3rd ed., Chap. 5, pp. 115–188). Sara Burgerhartstraat 25 P.O. Box 211, 1000 AE Amsterdam, The Netherlands: Elsevier Science B.V.

Stratoulias, V., Heino, T. I., & Michon, F. (2014). Lin-28 regulates oogenesis and muscle formation in drosophila melanogaster. *PLoS ONE*, *9*(6), e101141. doi:10.1371/journal.pone.0101141

Suominen, T., Uutela, P., Ketola, R. A., Bergquist, J., Hillered, L., Finel, M., ... Kostiainen, R. (2013). Determination of Serotonin and Dopamine Metabolites in Human Brain Microdialysis and Cerebrospinal Fluid Samples by UPLC-MS/MS: Discovery of Intact Glucuronide and Sulfate Conjugates. *PLoS ONE*, *8*(6), e68007. doi:10.1371/journal.pone.0068007

Supek, F., Bošnjak, M., Škunca, N., & Šmuc, T. (2011). REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. *PLoS ONE*, *6*(7), e21800. doi:10.1371/journal.pone.0021800

Sutherland, T. D., Unnithan, G. C., Andersen, J. F., Evans, P. H., Murataliev, M. B., Szabo, L. Z., ... Feyereisen, R. (1998). A cytochrome p450 terpenoid hydroxylase linked to the suppression of insect juvenile hormone synthesis. *Proceedings of the National Academy of Sciences*, *95*(22), 12884–12889. doi:10.1073/pnas.95.22.12884

Takeda, M. (1978). *Photoperiodic time measurement and seasonal adaptation of the southwestern corn borer, Diatraea grandiosella dyar (lepidoptera: Pyralidae)* (Doctoral dissertation, University of Missoury).

Takeda, M. & Skopik, S. D. (1997). Photoperiodic time measurement and related physiological mechanisms in insects and mites. *Annual Review of Entomology*, *42*(1), 323–349. doi:10.1146/annurev.ento.42.1.323

Tamaki, S., Takemoto, S., Uryu, O., Kamae, Y., & Tomioka, K. (2013). Opsins are involved in nymphal photoperiodic responses in the cricket *Modicogryllus siamensis*. *Physiological Entomology*, *38*(2), 163–172. doi:10.1111/phen.12015

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., & Kumar, S. (2011). MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsi-

mony methods. *Molecular Biology and Evolution*, *28*(10), 2731–2739. doi:10.1093/molbev/msr121

Tang, H. Y., Smith-Caldas, M. S., Driscoll, M. V., Salhadar, S., & Shingleton, A. W. (2011). FOXO regulates organ-specific phenotypic plasticity in Drosophila. *PLoS Genetics*, *7*(11). doi:10.1371/journal.pgen.1002373

Tatar, M. & Yin, C.-M. (2001). Slow aging during insect reproductive diapause: Why butterflies, grasshoppers and flies are like worms. *Experimental Gerontology*, *36*(4), 723–738. Slowly Aging Organisms. doi:https://doi.org/10.1016/S0531-5565(00)00238-2

Tauber, E. & Kyriacou, B. P. (2001). Insect photoperiodism and circadian clocks: Models and mechanisms. *Journal of Biological Rhythms*, *16*(4), 381–390. doi:10.1177/074873001129002088

Teerawanichpan, P., Robertson, A. J., & Qiu, X. (2010). A fatty acyl-CoA reductase highly expressed in the head of honey bee (apis mellifera) involves biosynthesis of a wide range of aliphatic fatty alcohols. *Insect Biochemistry and Molecular Biology*, *40*(9), 641–649. doi:10.1016/j.ibmb.2010.06.004

Teplova, M., Yuan, Y.-R., Phan, A. T., Malinina, L., Ilin, S., Teplov, A., & Patel, D. J. (2006). Structural basis for recognition and sequestration of UUU_{OH} 3' termini of nascent RNA polymerase III transcripts by la, a rheumatic disease autoantigen. *Molecular Cell*, *21*(1), 75–85. doi:10.1016/j.molcel.2005.10.027

Terakita, A. (2005). The opsins. *Genome Biology*, *6*(3), 213. doi:10.1186/gb-2005-6-3-213

Terakita, A., Kawano-Yamashita, E., & Koyanagi, M. (2011). Evolution and diversity of opsins. *Wiley Interdisciplinary Reviews: Membrane Transport and Signaling*, *1*(1), 104–111. doi:10.1002/wmts.6

The International Aphid Genomics Consortium. (2010). Genome sequence of the pea aphid *Acyrthosiphon pisum*. *PLOS Biology*, *8*(2), 1–24. doi:10.1371/journal.pbio.1000313

Thieme, T. & Dixon, A. F. G. (1996). Mate recognition in the aphis fabae complex: Daily rhythm of release and specificity of sex pheromones. *Entomologia Experimentalis et Applicata*, *79*(1), 85–89. doi:10.1111/j.1570-7458.1996.tb00812.x

Thomas, B. & Vince-Prue, D. (1996). *Photoperiodism in plants, second edition* (2nd ed.). Academic Press.

Tian, T., Liu, Y., Yan, H., You, Q., Yi, X., Du, Z., ... Su, Z. (2017). agriGO v2.0: a GO analysis toolkit for the agricultural community,

2017 update. *Nucleic Acids Research*, *45*(W1), W122–W129. doi:10.1093/nar/gkx382

Tournois, J. (1912). Influence de la lumiére sur la floraison du houblon japonais et du chanvre déterminées par des semis haitifs [Influence of light on flowering of japanese hops and hemp determined by hay seedlings]. *C.R. Acad. Sci. Paris*, *155*, 297–300.

Tournois, J. (1914). Etudes sur la sexualité du houblon [studies on the sexuality of hop]. *Annales des Sciences Naturelles (Botanique)*, *19*, 49–191.

Townson, S. M., Chang, B. S. W., Salcedo, E., Chadwell, L. V., Pierce, N. E., & Britt, S. G. (1998). Honeybee blue- and ultraviolet-sensitive opsins: Cloning, heterologous expression inDrosophila, and physiological characterization. *The Journal of Neuroscience*, *18*(7), 2412–2422. doi:10.1523/jneurosci.18-07-02412.1998

Trapnell, C., Hendrickson, D. G., Sauvageau, M., Goff, L., Rinn, J. L., & Pachter, L. (2012). Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nature Biotechnology*, *31*(1), 46–53. doi:10.1038/nbt.2450

Tsialikas, J. & Romer-Seibert, J. (2015). LIN28: Roles and regulation in development and beyond. *Development*, *142*(14), 2397–2404. doi:10.1242/dev.117580

Vafopoulou, X. & Steel, C. (2014). Synergistic induction of the clock protein period by insulin-like peptide and prothoracicotropic hormone in rhodnius prolixus (hemiptera): Implications for convergence of hormone signaling pathways. *Frontiers in physiology*, *5*, 41–41. doi:10.3389/fphys.2014.00041

van Doorn, W. G. (2003). Flower opening and closure: A review. *Journal of Experimental Botany*, *54*(389), 1801–1812. doi:10.1093/jxb/erg213

Vaz Nunes, M. & Saunders, D. (1999). Photoperiodic time measurement in insects: A review of clock models. *Journal of Biological Rhythms*, *14*(2), 84–104. doi:10.1177/074873099129000470

Vega, H., Waisfisz, Q., Gordillo, M., Sakai, N., Yanagihara, I., Yamada, M., … Joenje, H. (2005). Roberts syndrome is caused by mutations in ESCO2, a human homolog of yeast ECO1 that is essential for the establishment of sister chromatid cohesion. *Nature Genetics*, *37*(5), 468–470. doi:10.1038/ng1548

Velarde, R. A., Sauer, C. D., Walden, K. K. O., Fahrbach, S. E., & Robertson, H. M. (2005). Pteropsin: A vertebrate-like non-visual

opsin expressed in the honey bee brain. *Insect Biochemistry and Molecular Biology*, *35*(12), 1367–1377. doi:10.1016/j.ibmb.2005.09.001

Verlinden, H. (2018). Dopamine signalling in locusts and other insects. *Insect Biochemistry and Molecular Biology*, *97*, 40–52. doi:10.1016/j.ibmb.2018.04.005

Vieira, R., Míguez, J. M., & Aldegunde, M. (2005). GABA modulates day–night variation in melatonin levels in the cerebral ganglia of the damselfly ischnura graellsii and the grasshopper *Oedipoda caerulescens*. *Neuroscience Letters*, *376*(2), 111–115. doi:10.1016/j.neulet.2004.11.036

Vitaterna, M., King, D., Chang, A., Kornhauser, J., Lowrey, P., McDonald, J., ... Takahashi, J. (1994). Mutagenesis and mapping of a mouse gene, clock, essential for circadian behavior. *Science*, *264*(5159), 719–725. doi:10.1126/science.8171325

Vivien-Roels, B. & Pévet, P. (1993). Melatonin: Presence and formation in invertebrates. *Experientia*, *49*(8), 642–647. doi:10.1007/bf01923945

Walton, J. C., Weil, Z. M., & Nelson, R. J. (2011). Influence of photoperiod on hormones, behavior, and immune function. *Frontiers in Neuroendocrinology*, *32*(3), 303–319. doi:10.1016/j.yfrne.2010.12.003

Wang, B., Xiao, J.-H., Bian, S.-N., Niu, L.-M., Murphy, R. W., & Huang, D.-W. (2013). Evolution and expression plasticity of opsin genes in a fig pollinator, *Ceratosolen solmsi*. *PLoS ONE*, *8*(1), e53907. doi:10.1371/journal.pone.0053907

Wang, J.-L., Saha, T. T., Zhang, Y., Zhang, C., & Raikhel, A. S. (2017). Juvenile hormone and its receptor methoprene-tolerant promote ribosomal biogenesis and vitellogenesis in the *Aede aegypti* mosquito. *Journal of Biological Chemistry*, *292*(24), 10306–10315. doi:10.1074/jbc.m116.761387

Wang, Q., Egi, Y., Takeda, M., Oishi, K., & Sakamoto, K. (2015). Melatonin pathway transmits information to terminate pupal diapause in the Chinese oak silkmoth *Antheraea pernyi* and through reciprocated inhibition of dopamine pathway functions as a photoperiodic counter. *Entomological Science*, *18*(1), 74–84. doi:10.1111/ens.12083

Wang, Q., Hanatani, I., Takeda, M., Oishi, K., & Sakamoto, K. (2014). D2-like dopamine receptors mediate regulation of pupal diapause

in chinese oak silkmoth *Antheraea pernyi*. *Entomological Science*, *18*(2), 193–198. doi:10.1111/ens.12099

Wang, Q., Mohamed, A. A. M., & Takeda, M. (2013). Serotonin receptor b may lock the gate of PTTH release/synthesis in the chinese silk moth; antheraea pernyi: A diapause initiation/maintenance mechanism? *PLoS ONE*, *8*(11), e79381. doi:10.1371/journal.pone.0079381

Wang, T. (2005). Rhodopsin formation in drosophila is dependent on the PINTA retinoid-binding protein. *Journal of Neuroscience*, *25*(21), 5187–5194. doi:10.1523/jneurosci.0995-05.2005

Watanabe, T., Sadamoto, H., & Aonuma, H. (2011). Identification and expression analysis of the genes involved in serotonin biosynthesis and transduction in the field cricket *Gryllus bimaculatus*. *Insect Molecular Biology*, *20*(5), 619–635. doi:10.1111/j.1365-2583.2011.01093.x

Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., & Barton, G. J. (2009). Jalview version 2–a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, *25*(9), 1189–1191. doi:10.1093/bioinformatics/btp033

Williams, B. C., Garrett-Engele, C. M., Li, Z., Williams, E. V., Rosenman, E. D., & Goldberg, M. L. (2003). Two putative acetyltransferases, san and deco, are required for establishing sister chromatid cohesion in Drosophila. *Current biology : CB*, *13*(23), 2025–36. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/14653991

Wülbeck, C. & Helfrich-Förster, C. (2007). RNA in situ hybridizations on drosophila whole mounts. In *Methods in molecular biology* (pp. 495–511). Humana Press. doi:10.1007/978-1-59745-257-1_40

Yan, S., Zhu, J., Zhu, W., Zhang, X., Li, Z., Liu, X., & Zhang, Q. (2014). The expression of three opsin genes from the compound eye of *Helicoverpa armigera* (lepidoptera: Noctuidae) is regulated by a circadian clock, light conditions and nutritional status. *PLoS ONE*, *9*(10), e111683. doi:10.1371/journal.pone.0111683

Yang, L., Qin, Y., Li, X., Song, D., & Qi, M. (2007). Brain melatonin content and polyethism in adult workers of *Apis mellifera* and *Apis cerana* (hym., apidae). *Journal of Applied Entomology*, *131*(9-10), 734–739. doi:10.1111/j.1439-0418.2007.01229.x

Yasuo, S. & Yoshimura, T. (2009). Comparative analysis of the molecular basis of photoperiodic signal transduction in vertebrates. *Integrative and Comparative Biology*, *49*(5), 507–518. doi:10.1093/icb/icp011

Yin, Z.-J., Dong, X.-L., Kang, K., Chen, H., Dai, X.-Y., Wu, G.-A., ... Zhai, Y.-F. (2018). FoxO Transcription Factor Regulate Hormone Mediated Signaling on Nymphal Diapause. *Frontiers in Physiology*, *9*. doi:10.3389/fphys.2018.01654

Yoshitane, H., Ozaki, H., Terajima, H., Du, N.-H., Suzuki, Y., Fujimori, T., ... Fukada, Y. (2014). CLOCK-controlled polyphonic regulation of circadian rhythms through canonical and noncanonical e-boxes. *Molecular and Cellular Biology*, *34*(10), 1776–1787. doi:10.1128/mcb.01465-13

Yuan, Q., Metterville, D., Briscoe, A. D., & Reppert, S. M. (2007). Insect cryptochromes: Gene duplication and loss define diverse ways to construct insect circadian clocks. *Molecular Biology and Evolution*, *24*(4), 948–955. doi:10.1093/molbev/msm011

Zantke, J., Ishikawa-Fujiwara, T., Arboleda, E., Lohs, C., Schipany, K., Hallay, N., ... Tessmar-Raible, K. (2013). Circadian and circalunar clock interactions in a marine annelid. *Cell Reports*, *5*(1), 99–113. doi:10.1016/j.celrep.2013.08.031

Zera, A. J. (2016). Evolutionary endocrinology of hormonal rhythms: Juvenile hormone titer circadian polymorphism inGryllus firmus. *Integrative and Comparative Biology*, *56*(2), 159–170. doi:10.1093/icb/icw027

Zhang, Q. & Denlinger, D. (2011). Molecular structure of the prothoracicotropic hormone gene in the northern house mosquito, culex pipiens, and its expression analysis in association with diapause and blood feeding. *Insect molecular biology*, *20*(2), 201–213. doi:10.1111/j.1365-2583.2010.01058.x

Zhang, Y., Fan, J., Sun, J., Francis, F., & Chen, J. (2017). Transcriptome analysis of the salivary glands of the grain aphid, sitobion avenae. *Scientific Reports*, *7*(1). doi:10.1038/s41598-017-16092-z

Zhang, Wang, X.-X., Jing, X., Tian, H.-G., & Liu, T.-X. (2016). Winged Pea Aphids Can Modify Phototaxis in Different Development Stages to Assist Their Host Distribution. *Frontiers in Physiology*, *7*. doi:10.3389/fphys.2016.00307

Zhu, H., Sauman, I., Yuan, Q., Casselman, A., Emery-Le, M., Emery, P., & Reppert, S. M. (2008). Cryptochromes define a novel circadian clock mechanism in monarch butterflies that may underlie sun compass navigation. *PLoS Biology*, *6*(1), e4. doi:10.1371/journal.pbio.0060004

Zhu, H., Yuan, Q., Froy, O., Casselman, A., & Reppert, S. M. (2005). The two CRYs of the butterfly. *Current Biology*, *15*(23), R953–R954. doi:10.1016/j.cub.2005.11.030

Żyła, D., Homan, A., & Wegierek, P. (2017). Polyphyly of the extinct family oviparosiphidae and its implications for inferring aphid evolution (hemiptera, sternorrhyncha). *PLOS ONE*, *12*(4), e0174791. doi:10.1371/journal.pone.0174791

Appendix

Supplementary Tables and Figures

The Appendix section includes the headings describing the content of the tables referred to in the text and supplementary images. The full content of the tables listed below is accessible in the files provided in the attached CD.

Table S1: Differentially expressed genes when comparing aphids exposed to LD and SD_{14} conditions. Negative fold-change values represent downregulation in SD_{14}. Only genes with significant q-value are shown (adjusted p-value).

Table S2: Differentially expressed genes when comparing aphids exposed to LD and SD_{10} conditions. Negative fold-change values represent downregulation in SD_{10}. Only genes with significant q-value are shown (adjusted p-value).

Table S3: DAVID clustering tool results table. Clusters include differentially expressed genes (LD vs SD_{14} photoperiod) with similar annotations. The number of genes in each cluster is indicated.

Table S4: DAVID clustering tool results table. Clusters include differentially expressed genes (LD vs SD_{10} photoperiod) with similar annotations. The number of genes in each cluster is indicated.

Table S5: REVIGO results for enriched GO terms when comparing LD vs SD_{14} photoperiods

Table S6: REVIGO results for enriched GO terms when comparing LD vs SD_{10} photoperiods

Table S7: Differential expression data of opsin and cryptochrome genes in the transcriptomic experiment. LD vs SD_{14} photoperiods. Fold change: log2(fold change). q-value is an adjusted p-value considering false discovery rate (FDR). Empty slots (-) correspond to ACYPIs not detected in the transcriptome.

Table S8: Differential expression data of opsin and cryptochrome genes found in transcriptomic experiment. LD vs SD_{10}. Fold change: log2(fold change). Significant values in bold. q-value is an adjusted p-value considering false discovery rate (FDR).

Table S9: List of primers used for PCR amplification, sequencing, RT-qPCR and riboprobe synthesis of the genes characterised in the present report.

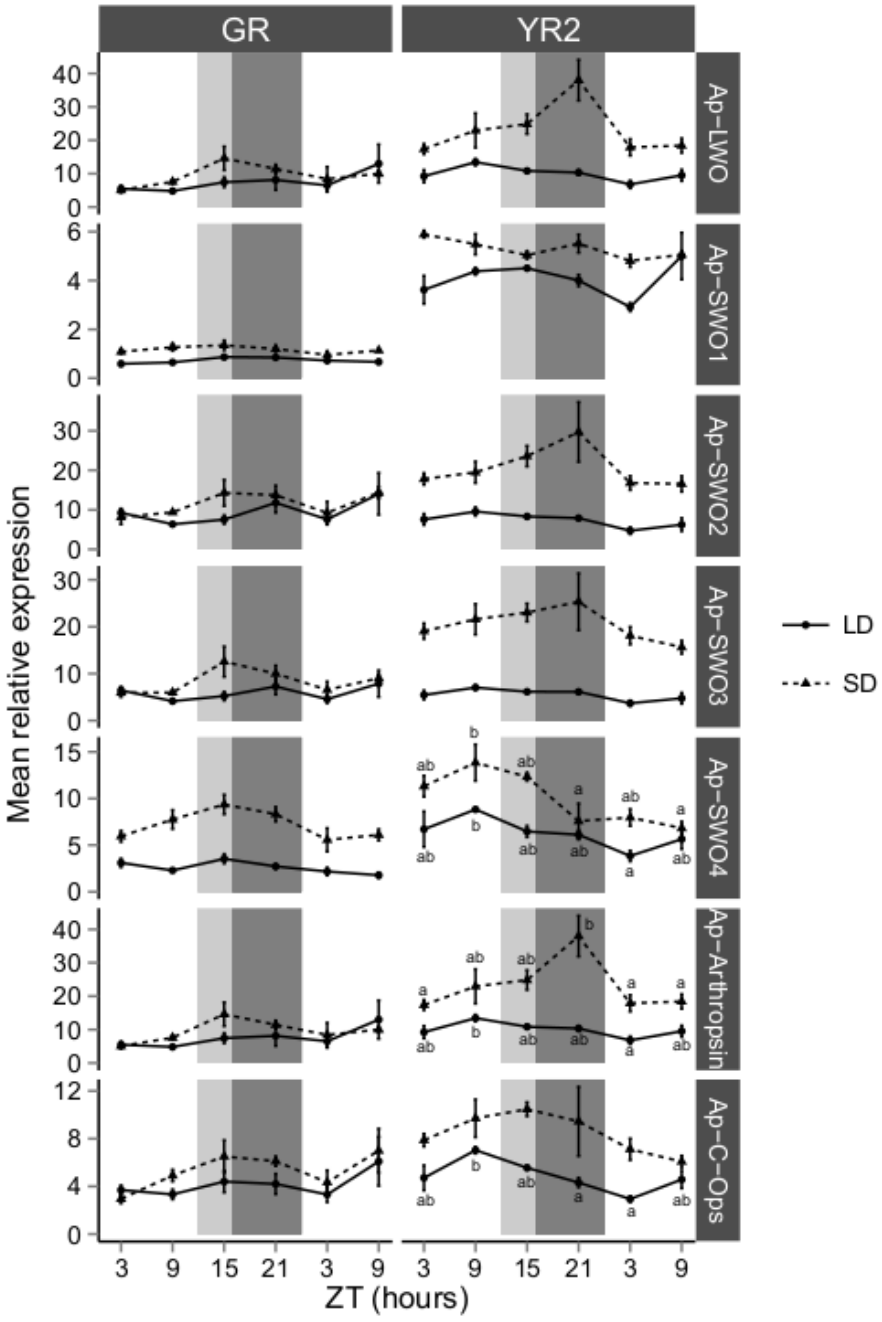

GR
YR2
Ap−LWO
Ap−SWO1
Ap−SWO2
Ap−SWO3
Ap−SWO4
Ap−Arthropsin
Ap−C−Ops
Mean relative expression
ZT (hours)
LD
SD

Figure S2: Previous page: Relative expression of opsin genes of *A. pisum* along 1,5 day/night cycles. Dark phases are shaded. Error bars represent the standard error of the mean (SEM). Homogeneous subgroups are labelled a; b; ab when significant differences were found (Tukey HSD, $P<0.05$). Ap-LWO, *A. pisum* Long Wavelength Opsin; Ap-SWO1 to Ap-SWO4, *A. pisum* Short Wavelength Opsins 1 to 4; Ap-Arthropsin, *A. pisum* Arthropsin; Ap-C-Ops, *A. pisum* C-opsin; GR, Gallur Red; YR2, York Red; ZT, zeitgeber; SD, Short day; LD, Long day.

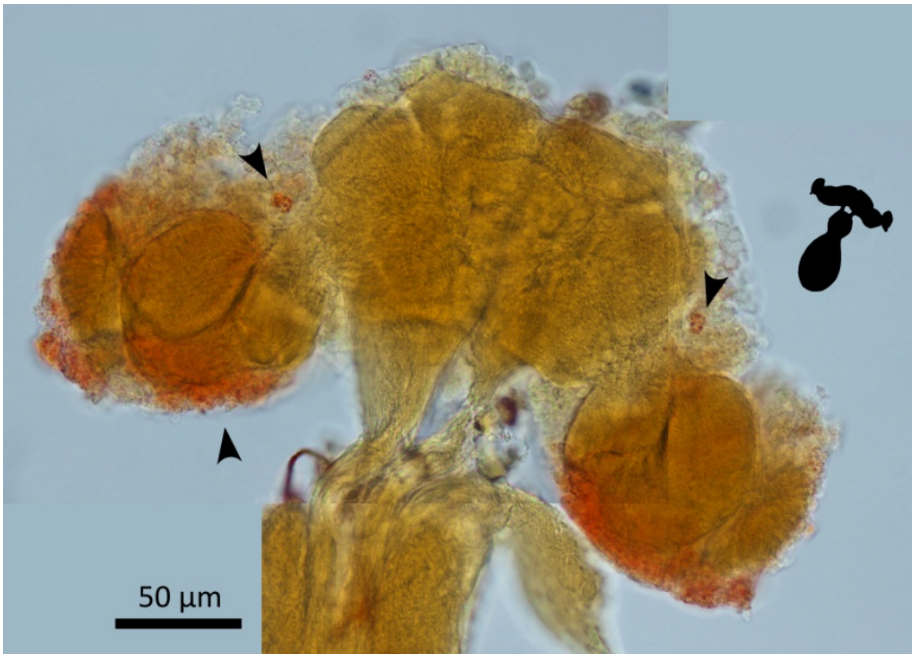


Figure S3: *In situ* localisation of Ap-Arthropsin gene transcripts in adult *A. pisum* central nervous system. Arrowheads point to specifically stained regions. A schematic representation of the aphid central nervous system is included beside each image indicating the orientation of the preparation. Riboprobes specifically bound are detected with Fast Red/HNPP (see Materials and Methods)

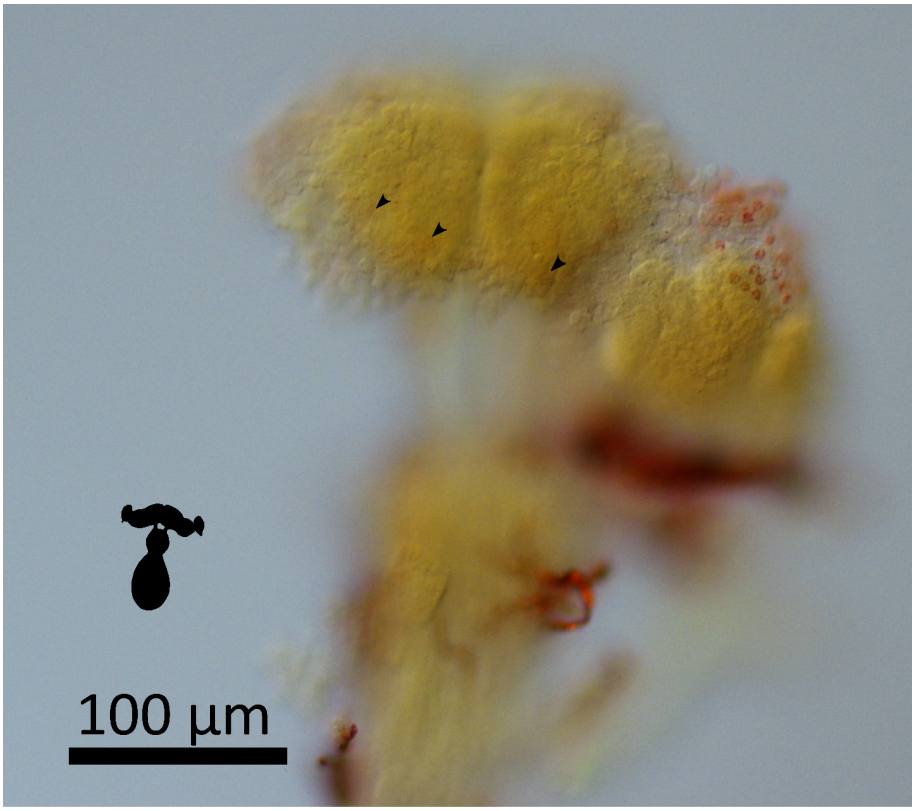


Figure S4: *In situ* localisation of Ap-C-ops gene transcripts in adult *A. pisum* central nervous system. Arrowheads point to specifically stained regions. A schematic representation of the aphid central nervous system is included beside each image indicating the orientation of the preparation. Riboprobes specifically bound are detected with Fast Red/HNPP (see Materials and Methods)

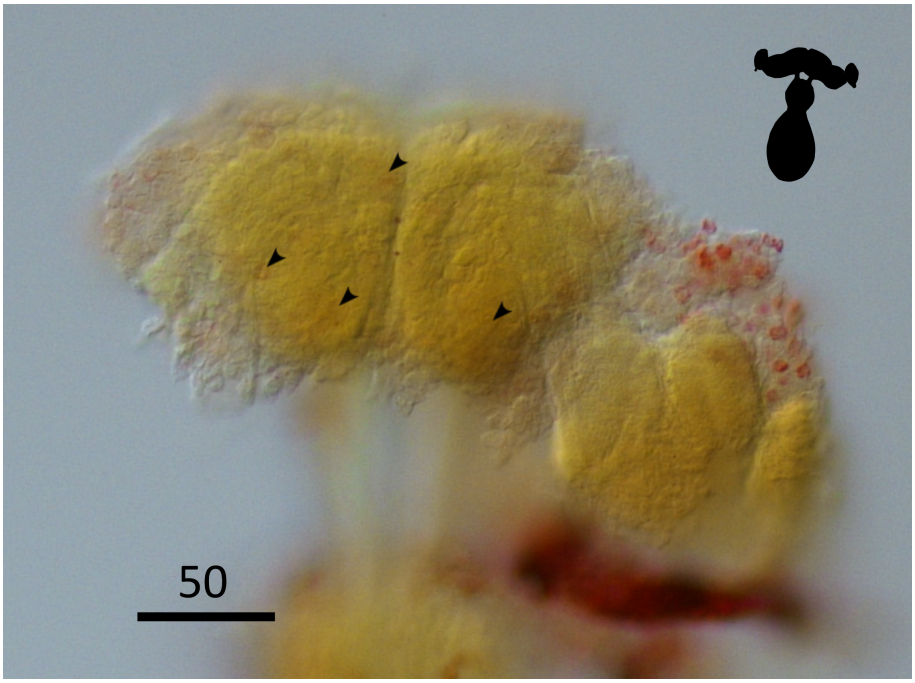


Figure S5: *In situ* localisation of Ap-C-ops gene transcripts in adult *A. pisum* central nervous system. This is an amplification of a different focus distance of Supplementary Figure S4. Arrowheads point to specifically stained regions. A schematic representation of the aphid central nervous system is included beside each image indicating the orientation of the preparation. Riboprobes specifically bound are detected with Fast Red/HNPP (see Materials and Methods)

Appendix

Supplementary Tables and Figures

Table S1: Differentially expressed genes when comparing aphids exposed to LD and SD_{14} conditions. Negative $\log_2$(fold change) values represent downregulation in SD_{14}. Only genes with significant q-value are shown (adjusted p-value).

ID	Gene	$\log_2$(fold change)	q-value
ACYPI000001	NW_003383490.1:LOC100144773	-1,89	0,01
ACYPI000002	NW_003383560.1:S1	-6,28	0
ACYPI000008	Cathepsin B	-3,59	0
ACYPI000009	Cathepsin B	-4,92	0
ACYPI000010	Cathepsin B	-3,38	0,03
ACYPI000019	Cathepsin B	-2,93	0
ACYPI000021	Cathepsin B	-3,47	0
ACYPI000093	ApisCSP5	-2,56	0
ACYPI000144	NW_003383568.1:LOC100158704	-2,38	0
ACYPI000220	NW_003384096.1:LOC100158786	1,96	0
ACYPI000291	cp21	-1,64	0,05
ACYPI000372	cp57	-2,99	0,01
ACYPI000394	NW_003384132.1:LOC100158979	-2,05	0
ACYPI000465	NW_003383885.1:LOC100159055	-1,72	0,01
ACYPI000473	NW_003383760.1:LOC100159064	-3,5	0

Continued on next page

Table S1 – continued from previous page

ID	Gene	$\log_2$(fold change)	q-value
ACYPI000499	NW_003383617.1:LOC100159096	-1,66	0
ACYPI000526	NW_003383583.1:LOC100159127	-2,33	0,03
ACYPI000536	NW_003383874.1:LOC100159138	-2,43	0,01
ACYPI000560	NW_003383726.1:LOC100159162	1,73	0
ACYPI000583	NW_003383575.1:LOC100159187	-1,12	0,04
ACYPI000631	AcypiEst11	-1,15	0,01
ACYPI000670	cp50	-1,64	0,02
ACYPI000807	NW_003384386.1:LOC100159435	-1,29	0,01
ACYPI000825	NW_003383553.1:LOC100159455	-1,46	0,03
ACYPI000889	cp62	-2,64	0
ACYPI000915	NW_003384254.1:LOC100159556	-1,97	0
ACYPI000961	NW_003384429.1:cp51	-3,65	0
ACYPI000983	cp18	-2,97	0
ACYPI001013	NW_003384756.1:LOC100159662	-3,09	0
ACYPI001032	NW_003384022.1:LOC100159682	-2,65	0
ACYPI001035	NW_003384333.1:LOC100159685	-8,05	0
ACYPI001084	cp54	-2,41	0,01
ACYPI001094	NW_003383682.1:LOC100159745	-4,89	0
ACYPI001144	NW_003384259.1:LOC100159797	-4,62	0
ACYPI001213	NW_003383816.1:LOC100159872	-3,68	0,05
ACYPI001245	NW_003384088.1:LOC100159905	-3,54	0
ACYPI001260	NW_003383512.1:LOC100159921	-2,79	0
ACYPI001340	NW_003384030.1:LOC100166807	-2,55	0
ACYPI001597	NW_003383524.1:LOC100160289	1,57	0,01
ACYPI001599	cp40	-2,91	0,01
ACYPI001610	NW_003383687.1:cprr1-10	-3,03	0,01
ACYPI001644	NW_003383529.1:cp44	-5,05	0
ACYPI001657	NW_003384475.1:LOC100160356	-5,67	0
ACYPI001681	NW_003383792.1:LOC100160381	-1,79	0,03
ACYPI001709	NW_003383525.1:LOC100160411	-3,33	0,03
ACYPI001780	NW_003383491.1:LOC100160486	-1,18	0,02
ACYPI001802	NW_003383562.1:LOC100160512	-2,51	0
ACYPI001863	NW_003383857.1:LOC100160576	-1,13	0,05
ACYPI001872	HMGCLL1	-1,69	0
ACYPI001904	NW_003383739.1:LOC100160620	1,49	0
ACYPI002017	NW_003384131.1:LOC100160745	-2,64	0
ACYPI002169	NW_003383522.1:LOC100160903	-3,32	0,02
ACYPI002182	Histone H2B	-3,04	0,01
ACYPI002185	NW_003383664.1:LOC100160919	-1,65	0
ACYPI002202	similar to maleless	-2,04	0,04
ACYPI002243	cp43	-5,27	0
ACYPI002246	NW_003384185.1:LOC100160987	1,64	0,02
ACYPI002250	Lipase13	-1,68	0,02
ACYPI002309	NW_003384341.1:LOC100161055	-3,04	0
ACYPI002311	ApisCSP10	-1,98	0,01
ACYPI002337	NW_003384362.1:LOC100161085	-2,17	0
ACYPI002419	NW_003384034.1:LOC100161176	1,5	0,01
ACYPI002494	NW_003384179.1:LOC100161259	-2,25	0
ACYPI002562	NW_003384429.1:LOC100161332	-2,03	0
ACYPI002564	NW_003384592.1:LOC100161334	-2,59	0,02
ACYPI002672	NW_003384857.1:LOC100161452	-2,53	0
ACYPI002751	NW_003383560.1:LOC100161535	-1,84	0
ACYPI002781	NW_003383716.1:LOC100161568	-2,63	0
ACYPI002788	NW_003387206.1:LOC100161576	-3,77	0

Continued on next page

Table S1 – continued from previous page

ID	Gene	$\log_2$(fold change)	q-value
ACYPI002830	NW_003383494.1:LOC100161621	-2,25	0,04
ACYPI002877	NW_003383747.1:cprr1-2	-1,18	0,03
ACYPI003067	AcypiEst12	-1,54	0,05
ACYPI003071	NW_003384058.1:LOC100161882	-1,61	0,01
ACYPI003134	Sp14	1,41	0
ACYPI003193	NW_003384893.1:LOC100168848	1,23	0,02
ACYPI003207	NW_003383495.1:LOC100162026	-2,08	0
ACYPI003229	cp52	-2,8	0
ACYPI003344	NW_003384196.1:LOC100162176	-1,73	0,03
ACYPI003395	NW_003384667.1:LOC100162231	-6,47	0
ACYPI003422	NW_003384732.1:LOC100162260	1,58	0,01
ACYPI003523	NW_003383923.1:LOC100162367	-2,16	0,01
ACYPI003561	NW_003383818.1:LOC100162409	-3,9	0
ACYPI003590	NW_003383594.1:LOC100162438	-1,54	0,02
ACYPI003597	NW_003383525.1:LOC100162445	-3,78	0
ACYPI003684	NW_003384307.1:LOC100162536	-2,71	0,01
ACYPI003692	NW_003383673.1:LOC100162544	-1,98	0,01
ACYPI003698	cp53	-2,41	0
ACYPI003827	NW_003383537.1:LOC100162688	-4,35	0,01
ACYPI003866	NW_003383787.1:LOC100162732	-2,55	0,01
ACYPI004114	NW_003384437.1:cp37	-3,35	0
ACYPI004280	NW_003383773.1:LOC100163179	-3,8	0
ACYPI004320	NW_003383874.1:LOC100163220	-3,18	0
ACYPI004413	NW_003384095.1:LOC100163319	1,42	0,03
ACYPI004418	NW_003383807.1:LOC100163324	1,86	0
ACYPI004446	NW_003383921.1:LOC100163352	-inf	0
ACYPI004810	NW_003384437.1:cp22	-1,68	0
ACYPI004819	NW_003383567.1:LOC100163759	-2,51	0
ACYPI004893	NW_003384437.1:cp34	-3,67	0,02
ACYPI004919	NW_003384259.1:LOC100163865	-2,02	0,02
ACYPI005002	NW_003383874.1:LOC100163954	-2,98	0,01
ACYPI005034	NW_003383730.1:LOC100163986	-1,16	0,05
ACYPI005037	NW_003383736.1:LOC100163990	-2,86	0
ACYPI005056	NW_003384371.1:LOC100164009	2,28	0,01
ACYPI005058	NW_003383584.1:LOC100164011	-1,04	0,05
ACYPI005070	NW_003383567.1:LOC100164026	1,43	0,02
ACYPI005113	P450 related to CYP4 family	-1,28	0,01
ACYPI005147	NW_003384009.1:LOC100164106	1,45	0,05
ACYPI005168	NW_003384149.1:LOC100164129	-2,55	0,03
ACYPI005252	NW_003383889.1:LOC100164217	-3,44	0
ACYPI005278	Lipase07	-1,75	0
ACYPI005339	NW_003384334.1:LOC100164310	-2,59	0
ACYPI005387	NW_003384600.1:cp10	-3,86	0,01
ACYPI005435	NW_003384085.1:LOC100164416	-5,06	0
ACYPI005549	NW_003383625.1:LOC100164540	-3,87	0,03
ACYPI005577	NW_003384048.1:LOC100164577	-1,77	0
ACYPI005624	NW_003385636.1:LOC100164630	-3,25	0
ACYPI005721	NW_003384056.1:LOC100164731	1,23	0,03
ACYPI005771	NW_003383666.1:LOC100164783	-2,71	0
ACYPI005919	NW_003383647.1:LOC100164941	1,72	0
ACYPI005985	NW_003383522.1:LOC100165011	-2,62	0
ACYPI006015	cp2	-2,11	0
ACYPI006045	m3AUGepi5s1111g148t1_gene	-5,83	0,01
ACYPI006052	NW_003383623.1:LOC100165083	1,43	0,02

Continued on next page

Table S1 – continued from previous page

ID	Gene	$\log_2$(fold change)	q-value
ACYPI006081	NW_003383593.1:LOC100165115	-4,21	0
ACYPI006175	cp76	-1,32	0,03
ACYPI006250	cprr1	-2,15	0
ACYPI006258	NW_003383874.1:LOC100165301	-1,33	0,01
ACYPI006259	NW_003384034.1:LOC100165302	1,67	0
ACYPI006265	NW_003384088.1:LOC100165308	-4,74	0
ACYPI006276	NW_003384096.1:LOC100165319	-2,3	0
ACYPI006282	NW_003383915.1:LOC100165325	-5,98	0,02
ACYPI006336	NW_003384122.1:LOC100165382	-2,35	0,02
ACYPI006529	NW_003383720.1:LOC100163510	-3,57	0
ACYPI006608	NW_003383560.1:LOC100165676	-2,54	0
ACYPI006655	NW_003383583.1:LOC100165726	-4,03	0
ACYPI006690	NW_003384529.1:LOC100165761	-4,04	0
ACYPI006791	NW_003384437.1:cp38	-3,39	0
ACYPI006825	NW_003383566.1:LOC100165911	-1,6	0
ACYPI006867	NW_003383527.1:LOC100165956	-1,94	0,01
ACYPI006872	NW_003383553.1:LOC100165962	-2,44	0
ACYPI006913	NW_003384367.1:LOC100166005	-1,64	0,02
ACYPI006969	NW_003383675.1:LOC100166065	-2,06	0
ACYPI006979	ACYPIP353756	-2,57	0
ACYPI007029	NW_003383503.1:LOC100166130	-2,01	0,03
ACYPI007067	NW_003383785.1:LOC100166173	2,32	0
ACYPI007077	NW_003383513.1:LOC100166184	-4,29	0
ACYPI007095	NW_003384334.1:LOC100166202	1,18	0,03
ACYPI007130	NW_003383736.1:LOC100166245	-2,26	0
ACYPI007152	NW_003383659.1:LOC100159033	-4,95	0
ACYPI007169	NW_003384259.1:LOC100166286	-2,02	0
ACYPI007284	NW_003383662.1:LOC100166409	-1,41	0,03
ACYPI007329	cp11	-2,79	0
ACYPI007430	NW_003384009.1:LOC100166571	-4,27	0
ACYPI007433	NW_003383871.1:LOC100166574	-5,84	0
ACYPI007463	NW_003383931.1:LOC100166605	-2,96	0
ACYPI007467	NW_003383560.1:LOC100160111	-3,44	0
ACYPI007575	NW_003383817.1:LOC100166724	-4,43	0
ACYPI007639	NW_003383604.1:LOC100166797	1,54	0,02
ACYPI007713	NW_003384113.1:LOC100166874	-1,17	0,02
ACYPI007753	NW_003383560.1:LOC100166916	-3,24	0
ACYPI007927	cp14	-3,21	0,01
ACYPI007931	NW_003383513.1:LOC100167113	-2,8	0
ACYPI008168	NW_003383773.1:LOC100167369	-3,01	0
ACYPI008179	NW_003383513.1:LOC100167380	-4,35	0,03
ACYPI008192	NW_003384108.1:LOC100167393	-1,38	0
ACYPI008199	NW_003383623.1:LOC100167400	1,54	0,01
ACYPI008261	NW_003383516.1:LOC100167464	1,24	0,05
ACYPI008270	AcypiEst5	-1,22	0,03
ACYPI008282	P450 (partial) related to CYP4a7	-4,7	0,01
ACYPI008289	NW_003384526.1:LOC100167494	-1,2	0,02
ACYPI008389	NW_003383659.1:LOC100167607	-2,51	0
ACYPI008413	juvenile horm. inducible protein 8	-3,71	0,02
ACYPI008432	NW_003383519.1:LOC100167655	-1,95	0
ACYPI008524	CP	-3,27	0
ACYPI008595	NW_003383637.1:LOC100167838	-2,29	0,01
ACYPI008608	NW_003383493.1:LOC100167851	-5,46	0
ACYPI008688	poly (ADP	-3,83	0

Continued on next page

Table S1 – continued from previous page

ID	Gene	$\log_2$(fold change)	q-value
ACYPI008726	NW_003384150.1:LOC100167984	-1,97	0
ACYPI008767	NW_003383706.1:LOC100168031	-1,82	0
ACYPI008876	ACYPIP968256	-2,29	0,01
ACYPI009024	NW_003384378.1:LOC100569852	-1,96	0,01
ACYPI009116	ApisCSP6	-1,79	0
ACYPI009171	NW_003383580.1:LOC100168475	2,89	0
ACYPI009176	NW_003383678.1:LOC100168481	-1,46	0,01
ACYPI009231	NW_003383991.1:LOC100168540	-2,51	0
ACYPI009247	NW_003383780.1:LOC100168557	-1,56	0,01
ACYPI009335	NW_003383673.1:LOC100168655	-1,36	0,03
ACYPI009352	NW_003383931.1:LOC100168672	-4,1	0
ACYPI009389	NW_003383991.1:LOC100168711	-2,22	0
ACYPI009456	NW_003384088.1:LOC100168782	-1,48	0,01
ACYPI009480	Oat	-2,68	0
ACYPI009525	NW_003383706.1:LOC100168857	-3,7	0
ACYPI009626	NW_003383506.1:LOC100168964	-2,31	0
ACYPI009637	NW_003383889.1:LOC100168978	-1,45	0,02
ACYPI009653	NW_003383527.1:LOC100168995	-3,91	0
ACYPI009713	NW_003384152.1:LOC100161712	1,33	0
ACYPI009803	cp15	-2,87	0
ACYPI009804	NW_003384437.1:cp36	-3,73	0
ACYPI009849	NW_003383535.1:LOC100169209	-5,92	0
ACYPI009940	NW_003384064.1:LOC100169311	-2,85	0
ACYPI009960	NW_003383948.1:LOC100169332	-2,13	0,01
ACYPI010031	NW_003383726.1:LOC100169409	1,2	0,05
ACYPI010055	ACYPIP949557	-2,81	0
ACYPI010057	NW_003383924.1:LOC100169440	-3,07	0
ACYPI010069	NW_003383760.1:LOC100169453	-1,65	0,03
ACYPI010088	NW_003384108.1:LOC100169473	-3,76	0,01
ACYPI010150	NW_003406627.1:LOC100169540	-2,52	0
ACYPI010188	NW_003384333.1:LOC100169582	2,18	0
ACYPI010224	NW_003383720.1:LOC100169622	2,19	0
ACYPI010230	NW_003384150.1:LOC100169629	-2,55	0
ACYPI060024	m3AUGpeaaphid0s472g16t1_gene	1,78	0
ACYPI060277	PHUM432760	inf	0,01
ACYPI060584	NW_003383635.1:LOC100164752	-4,68	0
ACYPI060630	NW_003383840.1:LOC100167957	1,19	0,05
ACYPI060701	NW_003383583.1:LOC100163851	-2,55	0
ACYPI060748	NW_003384333.1:LOC100570018	-8,24	0
ACYPI060829	sp25	-1,58	0,01
ACYPI060831	sp1	-1,41	0,01
ACYPI060832	cp9	-3,53	0,03
ACYPI060845	sp18	-1,42	0
ACYPI061253	m3AUGepi5s213g25t1_gene	-3,1	0,01
ACYPI061444	NW_003383613.1:LOC100302493	-3	0,01
ACYPI061541	NW_003383672.1:LOC100574390	-1,1	0,04
ACYPI061574	NW_003384303.1:LOC100570372	-4,61	0
ACYPI061710	cp38	-3,96	0
ACYPI062318	NW_003384007.1:LOC100568505	-2,02	0
ACYPI062383	NW_003383984.1:LOC100575796	-inf	0
ACYPI062389	NW_003383546.1:LOC100570619	1,51	0,02
ACYPI062462	NW_003383897.1:LOC100161473	-1,48	0,03
ACYPI062569	NW_003383675.1:LOC100575626	-5	0,01
ACYPI063324	NW_003384418.1:LOC100164574	-1,71	0,01

Continued on next page

Table S1 – continued from previous page

ID	Gene	$\log_2$(fold change)	q-value
ACYPI063338	NW_003383551.1:LOC100575653	-inf	0,03
ACYPI063340	NW_003385764.1:LOC100168312	-1,87	0,03
ACYPI063370	NW_003383707.1:LOC100164841	-5,06	0
ACYPI063385	NW_003384333.1:LOC100571146	-5,01	0
ACYPI063412	NW_003383667.1:LOC100569691	-6,35	0,01
ACYPI064209	NW_003397779.1:LOC100568654	-2,74	0
ACYPI064286	sp2	-2,13	0
ACYPI064288	cp17	-4	0,02
ACYPI064297	cp12	-4,14	0
ACYPI064944	NW_003383593.1:LOC100162725	-2,73	0,04
ACYPI065159	NW_003384914.1:LOC100569490	-2,57	0
ACYPI065224	cp7	-3,27	0
ACYPI065501	m3AUGepi4s3760g25t1_gene	-8,4	0
ACYPI065832	NW_003401411.1:LOC100572930	-inf	0
ACYPI065874	NW_003404974.1:LOC100575526	1,47	0,05
ACYPI066091	phantom	-1,77	0
ACYPI066784	NW_003384681.1:LOC100161636	-2,59	0
ACYPI066813	NW_003383573.1:LOC100575757	-4,38	0,03
ACYPI066831	NW_003383590.1:LOC100571349	-1,53	0
ACYPI066890	NW_003383501.1:LOC100575506	-1,77	0,03
ACYPI066895	NW_003383521.1:LOC100159543	1,22	0,03
ACYPI066937	NW_003384879.1:LOC100163145	-2,54	0,02
ACYPI066977	sp3	-1,88	0
ACYPI067456	m3AUGpeaaphid0s107g43t1_gene	-3,92	0
ACYPI067528	NW_003383546.1:LOC100569038	-3,45	0
ACYPI067812	NW_003384333.1:LOC100162717	1,54	0
ACYPI067836	NW_003383706.1:LOC100163825	-4,41	0,01
ACYPI067852	sp14	-1,47	0,01
ACYPI067864	sp12	-1,27	0,01
ACYPI068464	NW_003383591.1:LOC100162969	-5,85	0
ACYPI068652	NW_003383591.1:LOC100160947	-5,33	0
ACYPI068658	NW_003383596.1:LOC100570846	-4,26	0
ACYPI068863	m3AUGepir3s67g47t1_gene	-1,55	0
ACYPI069383	NW_003383635.1:LOC100162036	-1,87	0,01
ACYPI069462	NW_003383697.1:LOC100159715	-3,61	0
ACYPI069582	NW_003383672.1:LOC100574209	-1,34	0,01
ACYPI069646	ApisCSP13	-2,97	0,01
ACYPI070306	NW_003383785.1:LOC100164120	-4,34	0
ACYPI070406	NW_003387088.1:LOC100573727	-2,02	0,01
ACYPI070463	NW_003384119.1:LOC100568597	-3,26	0
ACYPI070469	NW_003384052.1:LOC100167307	1,1	0,04
ACYPI070504	cp33	-2,95	0
ACYPI070518	sp13	-1,4	0,01
ACYPI071113	NW_003397525.1:LOC100575469	-2,38	0,01
ACYPI071344	NW_003384600.1:LOC100160252	-3,81	0
ACYPI071387	cp28	-2,24	0,04
ACYPI071951	NW_003393720.1:LOC100168750	-1,63	0,03
ACYPI072004	NW_003384437.1:cp35	-4,18	0
ACYPI072111	NW_003384437.1:cp33	-3,31	0
ACYPI072242	sp10	-1,05	0,05
ACYPI072820	NW_003385387.1:LOC100574431	inf	0
ACYPI073031	cp5	-3,84	0,02
ACYPI073881	NW_003397865.1:LOC100166609	-3,38	0
ACYPI20268	NW_003383626.1:LOC100570295	1,8	0

Continued on next page

Table S1 – continued from previous page

ID	Gene	$\log_2$(fold change)	q-value
ACYPI21657	NW_003383633.1:LOC100570813	-3,26	0
ACYPI22397	NW_003383726.1:LOC100574338	-1,25	0,02
ACYPI24641	Retroviral envelope_13	-2,97	0,04
ACYPI26727	NW_003383706.1:LOC100574730	-3,11	0,01
ACYPI27105	NW_003383608.1:LOC100569251	2,18	0
ACYPI27192	NW_003383578.1:LOC100574876	-1,75	0,02
ACYPI29867	NW_003405103.1:LOC100575988	-4,1	0,02
ACYPI31715	NW_003383820.1:LOC100569360	-3,48	0
ACYPI32486	NW_003383664.1:LOC100575851	-1,59	0
ACYPI33613	NW_003383573.1:LOC100575965	-2,1	0,05
ACYPI34432	NW_003384334.1:LOC100571623	-2,91	0
ACYPI34702	NW_003383551.1:LOC100574260	-inf	0
ACYPI35003	NW_003384182.1:LOC100569882	-1,7	0,02
ACYPI35256	m3AUGpeaaphid0s154g13t1_gene	-1,22	0,02
ACYPI37215	NW_003383726.1:LOC100568588	1,75	0,01
ACYPI37842	NW_003383716.1:LOC100572094	-2,15	0,03
ACYPI38342	NW_003383522.1:LOC100163616	-2,09	0,01
ACYPI39140	NW_003383510.1:LOC100574721	-4,27	0
ACYPI39278	NW_003384119.1:LOC100568508	-2,33	0,02
ACYPI39361	NW_003383545.1:LOC100573621	-1,83	0
ACYPI42018	NW_003383617.1:LOC100575534	-1,41	0,01
ACYPI44101	NW_003383520.1:LOC100568885	-1,19	0,03
ACYPI44142	m3AUGepi4s736g74t1_gene	-2,49	0
ACYPI44373	NW_003383702.1:LOC100568636	-2,88	0
ACYPI45536	NW_003383845.1:LOC100568673	-2,54	0
ACYPI45645	NW_003384226.1:LOC100569884	-2,03	0,01
ACYPI46732	NW_003383892.1:LOC100570631	-1,85	0,03
ACYPI46913	NW_003384389.1:LOC100575671	-2,83	0,02
ACYPI47640	m3AUGepi4s126g34t1_gene	-3,59	0,03
ACYPI48365	NW_003383574.1:LOC100571126	-1,56	0,03
ACYPI48369	NW_003383605.1:LOC100570518	-1,65	0,01
ACYPI50492	NW_003384831.1:LOC100569341	-1,29	0,01
ACYPI50859	NW_003383591.1:LOC100575656	-2,96	0
ACYPI50967	NW_003384825.1:LOC100568811	1,18	0,02
ACYPI51884	NW_003383490.1:LOC100574300	-1,85	0,01
ACYPI52702	NW_003383502.1:Catb-16c	-3,1	0,03
ACYPI53411	NW_003383526.1:LOC100574750	-2,78	0
ACYPI54756	NW_003383835.1:LOC100569509	-1,78	0
ACYPI55916	NW_003383546.1:LOC100568915	-3,62	0
ACYPI56605	P450 (partial) related to CYP4 family	-2,75	0
ACYPI56617	NW_003400879.1:cp26	-4,66	0,02
ACYPI56619	cp23	-4,63	0
ACYPI56620	cp29	-2,46	0,02
ACYPI56624	cp70	-2,27	0,02

Table S2: Differentially expressed genes when comparing aphids exposed to LD and SD_{10} conditions. Negative log_2(fold change) values represent downregulation in SD_{10}. Only genes with significant q-value are shown (adjusted p-value).

ID	Gene	log_2(fold change)	q-value
ACYPI000001	NW_003383490.1:LOC100144773	-1,46	0,03
ACYPI000002	NW_003383560.1:S1	-3,95	0
ACYPI000008	Cathepsin B	-2,21	0
ACYPI000009	Cathepsin B	-3,86	0
ACYPI000010	Cathepsin B	-3,54	0
ACYPI000019	Cathepsin B	-3,36	0
ACYPI000021	Cathepsin B	-4,67	0
ACYPI000055	NW_003383721.1:Sec61a1	-1,28	0,01
ACYPI000063	mago nashi	-1,09	0,02
ACYPI000090	rpL24	-1,18	0,02
ACYPI000093	ApisCSP5	2,3	0,01
ACYPI000095	ApisCSP3	-2,21	0
ACYPI000097	ApisCSP4	3,42	0
ACYPI000108	NW_003384013.1:LOC100158668	-1,71	0
ACYPI000113	NW_003383715.1:LOC100158673	1,89	0
ACYPI000129	NW_003384023.1:LOC100158689	-0,98	0,04
ACYPI000132	ACYPIP455967	1,24	0,03
ACYPI000144	NW_003383568.1:LOC100158704	-1,19	0,04
ACYPI000159	NW_003383647.1:LOC100158720	-2,91	0
ACYPI000166	NW_003383554.1:LOC100158728	-2,33	0,01
ACYPI000174	NW_003383550.1:LOC100158736	1,1	0,02
ACYPI000176	P450 related to CYP4a7	2,29	0,02
ACYPI000202	NW_003383567.1:LOC100158767	0,99	0,04
ACYPI000219	NW_003383549.1:LOC100158785	-1,3	0
ACYPI000230	NW_003383894.1:LOC100158799	2,26	0
ACYPI000245	NW_003383553.1:LOC100158819	-1,57	0,01
ACYPI000257	NW_003383844.1:LOC100158833	-1,11	0,02
ACYPI000285	NW_003383763.1:LOC100158864	-2,36	0,01
ACYPI000297	NW_003383915.1:LOC100158876	0,99	0,04
ACYPI000319	NW_003383540.1:LOC100158899	-1,13	0,04
ACYPI000334	ApisOBP9	2,91	0
ACYPI000357	NW_003383582.1:LOC100158940	-1,7	0
ACYPI000364	NW_003383539.1:LOC100158947	1,07	0,04
ACYPI000375	NW_003383531.1:LOC100158958	-2,65	0
ACYPI000389	NW_003384258.1:LOC100158974	-3,27	0
ACYPI000394	NW_003384132.1:LOC100158979	-2,4	0,01
ACYPI000419	m3AUGepir3s158g32t1_gene	-1,6	0,04
ACYPI000433	NW_003383767.1:LOC100168016	-1,27	0
ACYPI000436	NW_003384219.1:LOC100159025	1,27	0,01
ACYPI000439	NW_003384237.1:LOC100159028	-1,8	0,02
ACYPI000451	NW_003384142.1:LOC100159040	-1,22	0,01
ACYPI000465	NW_003383885.1:LOC100159055	2,7	0,01
ACYPI000485	Similar to *D.melanogaster* CG3074	2,6	0
ACYPI000499	NW_003383617.1:LOC100159096	1,84	0
ACYPI000500	NW_003383846.1:Nop56	-1,17	0,04
ACYPI000506	NW_003383623.1:LOC100159104	-1,05	0,04
ACYPI000547	NW_003384055.1:LOC100159149	1,14	0,01
ACYPI000548	NW_003383706.1:LOC100159150	-1,79	0,02

Continued on next page

Table S2 – continued from previous page

ID	Gene	$\log_2$(fold change)	q-value
ACYPI000558	NW_003384367.1:LOC100159160	1,1	0,03
ACYPI000560	NW_003383726.1:LOC100159162	-1,16	0,04
ACYPI000571	NW_003383864.1:LOC100159173	-1,28	0,03
ACYPI000573	NW_003384318.1:LOC100159177	1,19	0,01
ACYPI000583	NW_003383575.1:LOC100159187	1,71	0
ACYPI000586	NW_003383698.1:LOC100159190	-2,46	0,01
ACYPI000594	NW_003383678.1:LOC100165000	-1,4	0,03
ACYPI000606	NW_003384248.1:LOC100159213	1,12	0,03
ACYPI000614	NW_003383560.1:LOC100159221	1,56	0
ACYPI000631	AcypiEst11	-1,17	0,01
ACYPI000655	NW_003383721.1:LOC100159264	1,96	0
ACYPI000677	NW_003384183.1:Det	-2,79	0,01
ACYPI000713	NW_003383496.1:LOC100159330	2,74	0
ACYPI000729	NW_003384060.1:LOC100159349	-1,43	0,03
ACYPI000751	NW_003383715.1:LOC100159374	-2	0
ACYPI000757	NW_003384580.1:LOC100159383	-1,34	0,01
ACYPI000802	NW_003384145.1:LOC100159429	2,29	0
ACYPI000807	NW_003384386.1:LOC100159435	1,66	0
ACYPI000814	NW_003383751.1:LOC100159444	-4,58	0
ACYPI000817	NW_003384303.1:LOC100570185	2,07	0
ACYPI000828	NW_003383570.1:LOC100159459	-1,28	0
ACYPI000835	NW_003383834.1:LOC100159466	-0,95	0,05
ACYPI000840	NW_003383726.1:LOC100159472	-3,28	0
ACYPI000865	NW_003383557.1:LOC100159501	1,16	0,01
ACYPI000867	NW_003383745.1:LOC100159503	-2,56	0
ACYPI000871	NW_003383542.1:LOC100159508	2,1	0
ACYPI000900	ltd2	1,67	0
ACYPI000906	NW_003383567.1:LOC100159546	-0,99	0,03
ACYPI000912	NW_003383626.1:LOC100159553	-1,19	0,01
ACYPI000919	NW_003384764.1:LOC100159560	-1,44	0,02
ACYPI000935	NW_003384637.1:LOC100159576	1,93	0,01
ACYPI000949	P450 related to CYP4a7	8,6	0,04
ACYPI000961	NW_003384429.1:cp51	1,33	0,01
ACYPI000965	NW_003384529.1:LOC100159608	-1,19	0,02
ACYPI000970	NW_003383760.1:LOC100159613	3,38	0
ACYPI000990	P450 related to CYP6ax1	1,65	0
ACYPI000991	NW_003383585.1:LOC100159637	1,87	0
ACYPI001007	NW_003383524.1:LOC100159654	-1,63	0,02
ACYPI001019	NW_003384474.1:LOC100159668	-2,62	0
ACYPI001034	putative ATIP_l	-3,81	0
ACYPI001035	NW_003384333.1:LOC100159685	2,4	0
ACYPI001037	NW_003384179.1:LOC100159687	-1,56	0,01
ACYPI001041	NW_003383510.1:LOC100159691	1,51	0
ACYPI001050	NW_003383828.1:LOC100159700	-3,14	0
ACYPI001094	NW_003383682.1:LOC100159745	-3,69	0
ACYPI001121	NW_003387424.1:LOC100159774	-1,1	0,02
ACYPI001127	NW_003383708.1:LOC100164289	1,73	0
ACYPI001150	NW_003383809.1:LOC100159803	-1,31	0,04
ACYPI001151	NW_003384566.1:LOC100159804	-2,43	0
ACYPI001175	cathepsin B	6,54	0
ACYPI001210	NW_003383918.1:LOC100159869	-1,39	0,01
ACYPI001245	NW_003384088.1:LOC100159905	1,4	0,02
ACYPI001297	NW_003383567.1:LOC100575509	-2,6	0
ACYPI001302	NW_003383596.1:LOC100159965	-2,66	0

Continued on next page

Table S2 – continued from previous page

ID	Gene	$\log_2$(fold change)	q-value
ACYPI001305	NW_003384893.1:LOC100159968	-6,02	0
ACYPI001351	NW_003383611.1:LOC100160017	-1,18	0,02
ACYPI001354	NW_003383593.1:LOC100160020	-1,8	0
ACYPI001382	NW_003384123.1:LOC100160050	-1,51	0,01
ACYPI001392	gustavus	1,19	0,02
ACYPI001402	NW_003383611.1:LOC100160074	-1,75	0
ACYPI001404	NW_003383598.1:LOC100160076	2,03	0,05
ACYPI001459	NW_003383518.1:LOC100160136	-1,1	0,02
ACYPI001471	NW_003383503.1:LOC100160149	1,08	0,04
ACYPI001505	NW_003384667.1:LOC100160190	1,32	0,04
ACYPI001519	P450 related to spook	-2,39	0,02
ACYPI001523	NW_003383838.1:LOC100168919	-1,13	0,02
ACYPI001535	NW_003384186.1:LOC100160220	-1,8	0
ACYPI001545	NW_003383560.1:Pla2g7	2,59	0
ACYPI001577	NW_003384459.1:LOC100160265	1,19	0,05
ACYPI001588	JH acid methyltransferase like	3,13	0,02
ACYPI001602	NW_003383793.1:LOC100160295	-1,09	0,03
ACYPI001641	NW_003384380.1:LOC100160338	-1,51	0,02
ACYPI001671	NW_003383499.1:LOC100160371	-1,14	0,04
ACYPI001673	NW_003383510.1:LOC100160373	-1,58	0
ACYPI001674	NW_003383750.1:LOC100160092	-3,11	0
ACYPI001677	NW_003384055.1:LOC100160377	2,05	0,01
ACYPI001712	NW_003383623.1:LOC100575876	-1,63	0,01
ACYPI001718	NW_003383625.1:LOC100160420	4,87	0
ACYPI001724	fbl	1,81	0
ACYPI001737	NW_003383577.1:LOC100160441	-1,09	0,04
ACYPI001761	NW_003383507.1:LOC100160466	1,28	0,01
ACYPI001771	NW_003383589.1:LOC100160476	-1,32	0,02
ACYPI001788	NW_003383795.1:LOC100165611	1,06	0,02
ACYPI001792	NW_003384510.1:LOC100160501	-3,31	0
ACYPI001796	NW_003384088.1:LOC100160505	-1,28	0
ACYPI001802	NW_003383562.1:LOC100160512	2,49	0,02
ACYPI001811	NW_003384495.1:LOC100160521	1,24	0,01
ACYPI001814	m3AUGepir3s436g17t1_gene	0,98	0,04
ACYPI001817	NW_003383568.1:LOC100160528	-0,96	0,04
ACYPI001820	NW_003383781.1:LOC100160531	-1,07	0,03
ACYPI001824	NW_003383520.1:LOC100160535	2,84	0
ACYPI001834	NW_003383567.1:LOC100160545	1,71	0,01
ACYPI001838	NW_003384037.1:LOC100160549	-1,01	0,03
ACYPI001843	NW_003384234.1:LOC100160554	1,47	0
ACYPI001859	NW_003383512.1:LOC100160572	-1,03	0,05
ACYPI001863	NW_003383857.1:LOC100160576	1,31	0
ACYPI001872	HMGCLL1	2,04	0,01
ACYPI001886	NW_003383976.1:LOC100160600	-2,32	0
ACYPI001887	NW_003383603.1:LOC100160601	1,57	0
ACYPI001901	NW_003383767.1:LOC100160617	1,08	0,04
ACYPI001904	NW_003383739.1:LOC100160620	-1,15	0,04
ACYPI001915	MELK	-1,59	0
ACYPI001926	CG7589	1,06	0,03
ACYPI001978	NW_003384689.1:LOC100160700	3,2	0
ACYPI001990	Spaetzle 6	-3,17	0
ACYPI002017	NW_003384131.1:LOC100160745	1,47	0,02
ACYPI002018	NW_003383724.1:LOC100160746	-1,56	0
ACYPI002043	NW_003383502.1:LOC100160772	-0,98	0,03

Continued on next page

Table S2 – continued from previous page

ID	Gene	$\log_2$(fold change)	q-value
ACYPI002049	NW_003384111.1:LOC100160778	-1,44	0
ACYPI002054	NW_003383592.1:LOC100160783	1,12	0,04
ACYPI002080	NW_003383560.1:LOC100160809	-1,48	0
ACYPI002086	NW_003384231.1:LOC100160815	-1,3	0,02
ACYPI002143	NW_003384022.1:LOC100160877	-1	0,05
ACYPI002160	NW_003383844.1:LOC100160894	-1,84	0
ACYPI002161	P450 related to CYP6ax1	1,43	0
ACYPI002164	similar to Viral A	-1,9	0
ACYPI002169	NW_003383522.1:LOC100160903	-3,22	0,03
ACYPI002172	CG11915	1,45	0,01
ACYPI002175	ACYPIP379951	2	0
ACYPI002182	Histone H2B	-2,9	0,02
ACYPI002185	NW_003383664.1:LOC100160919	3,32	0
ACYPI002199	NW_003384824.1:LOC100160933	-2,99	0
ACYPI002246	NW_003384185.1:LOC100160987	-1,99	0
ACYPI002250	Lipase13	2,08	0
ACYPI002255	Cyclin A	-2,63	0
ACYPI002268	NW_003383953.1:LOC100161010	-3,16	0
ACYPI002285	NW_003383497.1:LOC100161028	-1,08	0,02
ACYPI002293	NW_003383894.1:LOC100161037	-3,72	0
ACYPI002307	NW_003383761.1:LOC100161053	-1,01	0,05
ACYPI002311	ApisCSP10	2,02	0
ACYPI002337	NW_003384362.1:LOC100161085	-1,14	0,05
ACYPI002341	NW_003383946.1:LOC100161090	2,11	0,01
ACYPI002352	NW_003384289.1:LOC100161102	-2,26	0
ACYPI002361	NW_003383760.1:LOC100161111	-2,09	0
ACYPI002373	NW_003384149.1:LOC100161124	-2,03	0
ACYPI002408	NW_003383548.1:LOC100161165	-1,15	0,01
ACYPI002439	GPX	2,43	0
ACYPI002449	NW_003383864.1:LOC100161208	-1,22	0,03
ACYPI002459	NW_003383654.1:LOC100161219	1,56	0
ACYPI002460	NW_003383591.1:LOC100161220	1,64	0,04
ACYPI002481	Exportin	-1,02	0,04
ACYPI002494	NW_003384179.1:LOC100161259	-1,67	0
ACYPI002512	NW_003383629.1:LOC100161278	-2,24	0
ACYPI002513	NW_003383534.1:LOC100161279	1,08	0,03
ACYPI002543	NW_003383721.1:LOC100161311	3,85	0
ACYPI002562	NW_003384429.1:LOC100161332	-1,75	0,01
ACYPI002564	NW_003384592.1:LOC100161334	-2,24	0,02
ACYPI002567	NW_003383820.1:LOC100161337	-1	0,03
ACYPI002579	NW_003383948.1:LOC100161350	1,31	0,01
ACYPI002583	NW_003384387.1:LOC100161354	1,47	0,04
ACYPI002594	NW_003384469.1:LOC100161369	2,27	0
ACYPI002611	ACYPIP414366	1,5	0
ACYPI002622	NW_003383491.1:LOC100161399	-1,54	0,03
ACYPI002646	NW_003383944.1:LOC100161494	-2,16	0
ACYPI002649	NW_003383785.1:LOC100161429	-1,2	0,01
ACYPI002680	NW_003383736.1:LOC100161460	1,19	0,02
ACYPI002693	NW_003383726.1:LOC100161474	5,31	0
ACYPI002694	NW_003383802.1:LOC100570401	1,76	0,01
ACYPI002697	smoothened	-1,35	0
ACYPI002710	NW_003384587.1:LOC100161492	-2,79	0
ACYPI002746	NW_003383908.1:LOC100161530	1,52	0
ACYPI002751	NW_003383560.1:LOC100161535	5,04	0

Continued on next page

Table S2 – continued from previous page

ID	Gene	$\log_2$(fold change)	q-value
ACYPI002754	Tollo	-1,11	0,02
ACYPI002788	NW_003387206.1:LOC100161576	1,29	0,02
ACYPI002796	NW_003383788.1:LOC100161584	1,16	0,04
ACYPI002809	NW_003383504.1:LOC100161598	-1,09	0,02
ACYPI002810	NW_003384362.1:LOC100161599	1,24	0,01
ACYPI002813	NW_003384096.1:LOC100161602	-2,61	0
ACYPI002815	NW_003383525.1:LOC100161604	1,23	0,01
ACYPI002816	NW_003385550.1:LOC100161605	-3,21	0
ACYPI002825	NW_003383814.1:LOC100161615	1,3	0,02
ACYPI002830	NW_003383494.1:LOC100161621	-2,01	0
ACYPI002836	P450 related to CYP6 family	2,06	0
ACYPI002847	NW_003383572.1:LOC100161640	-1,01	0,03
ACYPI002848	NW_003384062.1:LOC100161641	-1,39	0,05
ACYPI002894	NW_003383799.1:LOC100161693	-1,29	0
ACYPI002905	NW_003383513.1:LOC100575560	-1,81	0
ACYPI002938	NW_003384258.1:LOC100161738	-1,22	0,01
ACYPI002976	NW_003383872.1:LOC100161779	1,55	0
ACYPI002993	NW_003384594.1:LOC100161796	-1,63	0
ACYPI002995	NW_003383595.1:LOC100574507	1,64	0
ACYPI002996	NW_003383513.1:LOC100161799	1,68	0
ACYPI003007	NW_003384202.1:LOC100161810	-1,04	0,05
ACYPI003008	NW_003383720.1:LOC100161811	1,22	0,01
ACYPI003023	NW_003383542.1:LOC100161829	1,11	0,05
ACYPI003035	FKBP45	-1,47	0
ACYPI003064	NW_003383750.1:LOC100161874	-1,46	0
ACYPI003071	NW_003384058.1:LOC100161882	-2,16	0
ACYPI003075	NW_003383693.1:LOC100161886	1,16	0,05
ACYPI003076	NW_003383842.1:LOC100161887	1,3	0,01
ACYPI003078	NW_003383767.1:LOC100161889	1,32	0,02
ACYPI003089	m3AUGepir3s288g87t1_gene	-1,63	0
ACYPI003102	NW_003383956.1:LOC100161914	-1,22	0
ACYPI003113	NW_003384415.1:LOC100161927	-4,15	0
ACYPI003123	NW_003383986.1:LOC100161937	-1,39	0,01
ACYPI003124	NW_003383501.1:LOC100161938	-1,82	0,01
ACYPI003128	NW_003383775.1:LOC100161942	1,58	0,01
ACYPI003130	NW_003384088.1:LOC100161944	-1,18	0,02
ACYPI003143	NW_003383726.1:LOC100161958	-1,39	0,02
ACYPI003175	NW_003384133.1:LOC100161992	1,72	0
ACYPI003193	NW_003384893.1:LOC100168848	-2,35	0
ACYPI003197	NW_003383508.1:LOC100162014	-2,27	0
ACYPI003200	NW_003383499.1:LOC100162018	1,3	0,02
ACYPI003206	NW_003384793.1:LOC100162025	-2,05	0
ACYPI003207	NW_003383495.1:LOC100162026	-2,11	0,01
ACYPI003208	NW_003384038.1:Nip7	-1,14	0,01
ACYPI003255	NW_003384149.1:LOC100162075	-2,26	0
ACYPI003269	NW_003383502.1:LOC100162092	1,17	0,04
ACYPI003295	NW_003383639.1:LOC100162121	-1,38	0,02
ACYPI003302	NW_003383576.1:LOC100167093	-2,55	0
ACYPI003313	NW_003383536.1:LOC100162140	4,57	0
ACYPI003317	NW_003383709.1:LOC100162145	-1,8	0
ACYPI003321	NW_003384259.1:LOC100162149	1,63	0,01
ACYPI003327	NW_003383889.1:LOC100162155	0,96	0,05
ACYPI003340	NW_003383765.1:LOC100162171	1,23	0,03
ACYPI003344	NW_003384196.1:LOC100162176	-1,74	0,01

Continued on next page

Table S2 – continued from previous page

ID	Gene	$\log_2$(fold change)	q-value
ACYPI003383	NW_003383708.1:LOC100162218	2,28	0,02
ACYPI003395	NW_003384667.1:LOC100162231	2,69	0,02
ACYPI003413	NW_003383986.1:LOC100162251	1,04	0,03
ACYPI003414	NW_003384105.1:LOC100162252	-3,32	0
ACYPI003435	NW_003384154.1:LOC100162275	-1,21	0,02
ACYPI003439	NW_003383499.1:LOC100162279	-0,97	0,04
ACYPI003450	NW_003384579.1:LOC100162290	2,43	0
ACYPI003481	NW_003383719.1:LOC100162322	-1,05	0,04
ACYPI003501	NW_003383814.1:LOC100162344	-1,13	0,02
ACYPI003520	Sp9	-3,82	0,01
ACYPI003523	NW_003383923.1:LOC100162367	-2,27	0,02
ACYPI003528	P450 related to CYP6ax1	1,6	0
ACYPI003541	Ppif	-1,4	0,03
ACYPI003561	NW_003383818.1:LOC100162409	-2,28	0
ACYPI003574	NW_003384753.1:LOC100162422	-1,11	0,01
ACYPI003579	NW_003383612.1:LOC100162427	1,2	0,01
ACYPI003590	NW_003383594.1:LOC100162438	2,56	0
ACYPI003597	NW_003383525.1:LOC100162445	-3,42	0
ACYPI003629	Cyclin	-1,8	0,01
ACYPI003638	NW_003387424.1:LOC100570083	-1,65	0,01
ACYPI003644	NW_003383569.1:LOC100162494	1,27	0,01
ACYPI003647	NW_003383524.1:LOC100162497	1,66	0
ACYPI003668	snf	-1,02	0,02
ACYPI003678	NW_003384835.1:LOC100162529	1,07	0,03
ACYPI003692	NW_003383673.1:LOC100162544	2,17	0
ACYPI003719	NW_003383835.1:LOC100162573	1,81	0,01
ACYPI003765	NW_003384309.1:LOC100162620	-3,02	0
ACYPI003827	NW_003383537.1:LOC100162688	-2,21	0,03
ACYPI003831	NW_003383573.1:LOC100162692	-1,15	0,04
ACYPI003857	NW_003384525.1:LOC100162722	1,37	0,01
ACYPI003873	NW_003383622.1:LOC100162741	-2,67	0
ACYPI003894	AcypiDlp5	-2,87	0,02
ACYPI003916	Sp4	-5,08	0
ACYPI003917	NW_003384486.1:LOC100162791	1,59	0,04
ACYPI003936	NW_003383513.1:LOC100162813	-1,08	0,01
ACYPI003950	NW_003384046.1:PDE11A	1,07	0,02
ACYPI003971	NW_003384209.1:LOC100162848	3,38	0
ACYPI003979	NW_003385051.1:LOC100162856	2,11	0,01
ACYPI003988	NW_003383501.1:LOC100162866	-1,5	0,02
ACYPI003995	NW_003383825.1:LOC100162873	1,76	0
ACYPI003997	NW_003384096.1:LOC100162875	-1,79	0
ACYPI004037	hephaestus	-1,51	0
ACYPI004039	NW_003383560.1:LOC100162920	-1,06	0,05
ACYPI004055	NW_003383529.1:LOC100162936	-1,02	0,03
ACYPI004063	NW_003383922.1:LOC100162944	-2,65	0
ACYPI004090	NW_003383747.1:LOC100162972	-1,45	0,01
ACYPI004091	NW_003383635.1:LOC100162973	1,36	0,01
ACYPI004092	NW_003384479.1:LOC100162974	-2,69	0
ACYPI004128	NW_003384727.1:LOC100163012	-1,2	0,05
ACYPI004148	NW_003383593.1:LOC100163035	2,21	0
ACYPI004152	NW_003383504.1:Mad2	-1,52	0,01
ACYPI004165	NW_003383528.1:LOC100163053	-4,28	0
ACYPI004176	NW_003383603.1:LOC100163065	1,15	0,02
ACYPI004217	m3AUGepir2s134g6t1_gene	1,55	0,01

Continued on next page

Table S2 – continued from previous page

ID	Gene	$\log_2$(fold change)	q-value
ACYPI004243	NW_003383551.1:LOC100163139	-3,47	0,02
ACYPI004274	NW_003383882.1:LOC100574977	-1,13	0,01
ACYPI004298	NW_003383838.1:LOC100163198	1,71	0
ACYPI004320	NW_003383874.1:LOC100163220	-2,33	0,03
ACYPI004325	NW_003384344.1:LOC100163225	2,39	0,01
ACYPI004328	NW_003383894.1:LOC100163228	-1,09	0,02
ACYPI004341	NW_003383773.1:LOC100163242	-3,47	0
ACYPI004344	NW_003383530.1:LOC100163245	-2,06	0,03
ACYPI004351	NW_003384296.1:Nap1l1	-1,07	0,03
ACYPI004357	NW_003383516.1:LOC100163260	-1,16	0,02
ACYPI004363	NW_003383781.1:LOC100163266	-1,72	0
ACYPI004381	NW_003383781.1:LOC100163285	-2,22	0,01
ACYPI004403	M1 metalloprotease	1,2	0,02
ACYPI004408	Ldlrap1	1,02	0,02
ACYPI004466	NW_003384663.1:LOC100163373	-1,16	0,03
ACYPI004475	NW_003383789.1:LOC100165440	-1,23	0,03
ACYPI004484	NW_003384639.1:LOC100163393	-1,77	0,02
ACYPI004502	NW_003384365.1:LOC100569675	-0,99	0,03
ACYPI004516	Serine Protease Inhibitor 5	1,22	0,03
ACYPI004517	NW_003383963.1:LOC100163428	-1,85	0
ACYPI004545	Sp7	-1,5	0
ACYPI004558	NW_003384603.1:LOC100163469	-1,53	0
ACYPI004630	NW_003383709.1:LOC100163550	-1,01	0,03
ACYPI004647	NW_003384578.1:LOC100163569	2	0
ACYPI004684	NW_003383756.1:LOC100163609	-1,41	0,04
ACYPI004691	NW_003383984.1:LOC100163617	-2,64	0
ACYPI004701	NW_003383869.1:LOC100163628	1,15	0,01
ACYPI004702	NW_003383551.1:LOC100163629	0,97	0,03
ACYPI004735	NW_003383752.1:LOC100163667	-1,57	0
ACYPI004741	NW_003383698.1:LOC100163673	-2,1	0
ACYPI004807	NW_003383915.1:LOC100163746	6,75	0
ACYPI004810	NW_003384437.1:cp22	1,59	0,01
ACYPI004819	NW_003383567.1:LOC100163759	-2,11	0,01
ACYPI004824	ACYPIP106532	-2,37	0
ACYPI004850	NW_003383514.1:LOC100163792	-1,94	0
ACYPI004865	NW_003384108.1:LOC100163808	-1,37	0
ACYPI004891	CypB	-1,19	0,04
ACYPI004919	NW_003384259.1:LOC100163865	1,83	0
ACYPI004929	NW_003383935.1:LOC100163875	-1,38	0
ACYPI004954	NW_003383577.1:LOC100163902	-1,07	0,03
ACYPI004973	NW_003383927.1:LOC100163923	1,14	0,01
ACYPI004977	NW_003383515.1:LOC100163927	-1,08	0,02
ACYPI005009	NW_003383944.1:LOC100163961	1,11	0,03
ACYPI005012	NW_003383500.1:LOC100163964	1,46	0,03
ACYPI005035	NW_003383497.1:LOC100163987	-1,31	0,02
ACYPI005042	NW_003383952.1:LOC100163995	1,73	0,05
ACYPI005048	l(2)gl	-1,11	0,02
ACYPI005053	NW_003385082.1:LOC100164006	2,29	0
ACYPI005057	NW_003383590.1:LOC100164010	-1,45	0
ACYPI005058	NW_003383584.1:LOC100164011	1,3	0,02
ACYPI005059	NW_003384362.1:LOC100164012	-4,52	0
ACYPI005068	NW_003383603.1:LOC100164023	-2,13	0
ACYPI005072	NW_003385903.1:LOC100164028	1,03	0,04
ACYPI005106	DP2	-1,18	0,03

Continued on next page

Table S2 – continued from previous page

ID	Gene	$\log_2$(fold change)	q-value
ACYPI005107	NW_003383633.1:LOC100164066	1,51	0
ACYPI005112	NW_003383940.1:LOC100164071	-1,03	0,04
ACYPI005141	m3AUGepir3s844g81t1_gene	-3,74	0
ACYPI005155	NW_003384056.1:LOC100164114	-1,21	0,01
ACYPI005156	NW_003383923.1:LOC100164115	1,62	0
ACYPI005157	NW_003384276.1:LOC100164117	-1,28	0,01
ACYPI005179	NW_003383548.1:LOC100574964	-2,6	0
ACYPI005191	NW_003384225.1:LOC100164153	-1,73	0,01
ACYPI005212	Sp3	-2,67	0
ACYPI005229	NW_003383558.1:LOC100164193	-1,18	0,01
ACYPI005238	NW_003383726.1:LOC100164203	-1,65	0,04
ACYPI005252	NW_003383889.1:LOC100164217	-2,96	0
ACYPI005270	NW_003383677.1:LOC100164236	-1,75	0
ACYPI005273	NW_003383614.1:Myo20	-2,05	0
ACYPI005278	Lipase07	1,87	0
ACYPI005284	NW_003383518.1:LOC100164253	-1,32	0,01
ACYPI005322	NW_003383625.1:LOC100164293	-2,07	0
ACYPI005329	NW_003384970.1:LOC100164300	1,09	0,04
ACYPI005336	NW_003383581.1:LOC100164307	-1,96	0
ACYPI005337	m3AUGepir2p1s17g102t1_gene	2,56	0
ACYPI005339	NW_003384334.1:LOC100164310	-1,5	0,03
ACYPI005340	Histone H2A.Z/F	-1,04	0,03
ACYPI005344	NW_003384531.1:LOC100164316	-1,73	0,03
ACYPI005349	NW_003384068.1:LOC100164321	-1,26	0
ACYPI005371	NW_003383610.1:LOC100164343	-3,03	0
ACYPI005384	ApisOBP5	-2,28	0
ACYPI005398	NW_003383736.1:LOC100164376	-2,24	0
ACYPI005411	NW_003383527.1:LOC100164389	2,25	0
ACYPI005466	NW_003385794.1:LOC100164447	-1,41	0
ACYPI005489	NW_003384649.1:LOC100164472	1,23	0,02
ACYPI005495	NW_003384022.1:LOC100164480	-1,28	0,01
ACYPI005525	NW_003384753.1:LOC100164514	-1,4	0
ACYPI005530	NW_003383721.1:LOC100164519	-1,23	0,01
ACYPI005534	NW_003383846.1:LOC100164524	-0,96	0,04
ACYPI005546	NW_003383833.1:LOC100164537	-1,15	0,02
ACYPI005549	NW_003383625.1:LOC100164540	-3,05	0
ACYPI005573	NW_003383507.1:LOC100164570	2,48	0,05
ACYPI005582	NW_003383760.1:LOC100570692	1,73	0,01
ACYPI005590	NW_003383524.1:LOC100164594	-3,72	0
ACYPI005594	NW_003383656.1:LOC100164598	-1,29	0,03
ACYPI005596	P450 related to CYP301a1	-1,03	0,03
ACYPI005624	NW_003385636.1:LOC100164630	-1,79	0
ACYPI005643	NW_003383805.1:LOC100164650	1,52	0
ACYPI005644	NW_003384395.1:LOC100164651	-2,72	0
ACYPI005681	NW_003383520.1:LOC100164691	2,02	0
ACYPI005699	NW_003383490.1:LOC100164709	2,04	0
ACYPI005706	NW_003383783.1:LOC100164716	1,13	0,02
ACYPI005731	ACYPIP382066	-0,96	0,04
ACYPI005732	NW_003383887.1:LOC100164742	1,16	0,03
ACYPI005755	NW_003383501.1:LOC100162473	-1,01	0,04
ACYPI005765	NW_003383573.1:LOC100164777	-3,78	0
ACYPI005769	NW_003383491.1:LOC100164781	1,06	0,03
ACYPI005787	NW_003384396.1:Nop58	-1,86	0
ACYPI005808	Rab23	-0,96	0,05

Continued on next page

Table S2 – continued from previous page

ID	Gene	$\log_2$(fold change)	q-value
ACYPI005816	NW_003383491.1:LOC100164829	-2,58	0,01
ACYPI005818	NW_003384383.1:LOC100164831	1,85	0
ACYPI005838	NW_003383847.1:LOC100164853	1,42	0
ACYPI005839	NW_003383651.1:LOC100164854	-1,06	0,02
ACYPI005871	Cyclin D	-1,96	0,03
ACYPI005902	NW_003383963.1:LOC100164922	-1,96	0
ACYPI005910	NW_003383887.1:LOC100167115	2,4	0
ACYPI005959	NW_003384022.1:LOC100164984	-1,61	0,01
ACYPI005960	NW_003383523.1:LOC100164985	1,2	0,02
ACYPI005983	NW_003384793.1:LOC100165009	1,84	0
ACYPI005985	NW_003383522.1:LOC100165011	-1,6	0,03
ACYPI006048	NW_003383915.1:LOC100165079	-3,48	0
ACYPI006069	NW_003383590.1:LOC100165102	-0,96	0,04
ACYPI006081	NW_003383593.1:LOC100165115	-3,98	0
ACYPI006094	NW_003384041.1:LOC100165130	-1,59	0
ACYPI006098	NW_003384198.1:LOC100165135	-1,23	0,01
ACYPI006115	NW_003383615.1:LOC100165153	-3,16	0
ACYPI006148	NW_003383549.1:LOC100165187	2,29	0
ACYPI006157	NW_003384059.1:LOC100165196	-2,12	0
ACYPI006183	NW_003384074.1:LOC100165223	-1,74	0,03
ACYPI006185	NW_003383567.1:LOC100165225	-1,16	0,01
ACYPI006190	NW_003384400.1:LOC100165230	-2,69	0
ACYPI006202	NW_003383844.1:LOC100165242	-3,73	0
ACYPI006207	ACYPIP861107	1,36	0,03
ACYPI006216	NW_003383495.1:LOC100165258	1,3	0,02
ACYPI006221	NW_003383824.1:LOC100165263	-1,55	0
ACYPI006230	NW_003383647.1:LOC100166788	-1,89	0,01
ACYPI006250	cprr1	2,27	0
ACYPI006256	Histone methyltransferase (SUV3	-0,99	0,05
ACYPI006258	NW_003383874.1:LOC100165301	1,2	0,03
ACYPI006259	NW_003384034.1:LOC100165302	-1,5	0,01
ACYPI006264	NW_003384835.1:LOC100165307	-3,05	0
ACYPI006265	NW_003384088.1:LOC100165308	2,37	0
ACYPI006266	NW_003383585.1:LOC100165309	-1,11	0,03
ACYPI006282	NW_003383915.1:LOC100165325	2,64	0
ACYPI006300	NW_003384350.1:LOC100165346	1,07	0,03
ACYPI006308	NW_003383603.1:LOC100165354	1,24	0,01
ACYPI006309	NW_003384018.1:LOC100165355	-1,19	0,01
ACYPI006312	NW_003383672.1:LOC100165358	1,5	0,05
ACYPI006328	NW_003383815.1:LOC100165374	1,36	0,04
ACYPI006355	NW_003384488.1:LOC100165402	-4,87	0
ACYPI006357	NW_003384046.1:LOC100165404	1,92	0
ACYPI006362	poly (ADP	-1,05	0,02
ACYPI006363	NW_003383546.1:LOC100165410	-1,1	0,02
ACYPI006367	NW_003384154.1:LOC100165415	-1,03	0,03
ACYPI006368	NW_003383992.1:LOC100165416	-1,34	0,05
ACYPI006371	NW_003383684.1:LOC100165419	-1,23	0,02
ACYPI006374	NW_003383766.1:LOC100165422	-2,57	0
ACYPI006387	NW_003383975.1:Aqp1	1,31	0
ACYPI006398	NW_003383948.1:LOC100165447	-1,93	0
ACYPI006409	NW_003383503.1:LOC100165458	-1,33	0,01
ACYPI006442	NW_003383523.1:LOC100165493	-2,33	0
ACYPI006449	NW_003383506.1:LOC100165501	-1,64	0,04
ACYPI006450	NW_003384138.1:LOC100165503	1,37	0,02

Continued on next page

Table S2 – continued from previous page

ID	Gene	$\log_2$(fold change)	q-value
ACYPI006464	NW_003383498.1:LOC100165519	1,09	0,04
ACYPI006491	NW_003383868.1:LOC100165551	-1,1	0,03
ACYPI006495	NW_003383700.1:Obp6	0,98	0,05
ACYPI006499	NW_003383566.1:LOC100165559	-2,59	0
ACYPI006520	NW_003383817.1:LOC100165584	1,35	0
ACYPI006523	NW_003384047.1:LOC100165587	-3,85	0
ACYPI006529	NW_003383720.1:LOC100163510	-1,97	0
ACYPI006571	NW_003384680.1:LOC100165637	1,39	0,02
ACYPI006583	NW_003383527.1:LOC100165651	-1,46	0
ACYPI006598	ACYPIP702646	1,39	0,04
ACYPI006608	NW_003383560.1:LOC100165676	1,67	0
ACYPI006651	NW_003384022.1:LOC100165722	-2,3	0,02
ACYPI006655	NW_003383583.1:LOC100165726	-2,56	0
ACYPI006664	3	-1,49	0
ACYPI006681	NW_003384085.1:LOC100165752	2,1	0
ACYPI006728	NW_003383664.1:LOC100165805	-1,16	0,02
ACYPI006729	P450 related to disembodied	-1,75	0,02
ACYPI006734	NW_003383809.1:LOC100165811	-1,19	0,01
ACYPI006735	NW_003383524.1:LOC100165812	1,35	0,02
ACYPI006737	NW_003384099.1:LOC100165814	-1,11	0,01
ACYPI006744	NW_003383493.1:LOC100165822	1,25	0,02
ACYPI006777	PRMT related Histone methyltransf.	-1,1	0,01
ACYPI006784	ilvA	2,22	0
ACYPI006801	NW_003383838.1:LOC100165884	-1,47	0
ACYPI006805	NW_003383792.1:LOC100165889	-1,86	0
ACYPI006811	Spaetzle 4	-1,99	0
ACYPI006820	Cactus	1,34	0
ACYPI006825	NW_003383566.1:LOC100165911	1,52	0
ACYPI006835	NW_003383603.1:LOC100165921	-4,42	0
ACYPI006882	P450 related to CYP6ax1	1,55	0,01
ACYPI006902	NW_003384100.1:LOC100165993	-1,02	0,02
ACYPI006918	NW_003384221.1:LOC100166010	1,19	0,01
ACYPI006924	NW_003384454.1:MDN1	-1	0,03
ACYPI006926	NW_003383864.1:LOC100575699	1,59	0
ACYPI006931	Histone H1	-inf	0,04
ACYPI006969	NW_003383675.1:LOC100166065	1,33	0,04
ACYPI006979	ACYPIP353756	-2,23	0
ACYPI006991	NW_003383855.1:LOC100166090	-1,35	0,01
ACYPI007024	NW_003383490.1:LOC100166125	-2,44	0,01
ACYPI007057	NW_003383810.1:LOC100166159	1,1	0,02
ACYPI007064	NW_003383590.1:LOC100166169	-1,8	0,02
ACYPI007077	NW_003383513.1:LOC100166184	-3,35	0
ACYPI007095	NW_003384334.1:LOC100166202	-2,44	0
ACYPI007126	NW_003383772.1:LOC100166240	-3,38	0
ACYPI007134	NW_003384750.1:LOC100166249	-1,24	0,01
ACYPI007152	NW_003383659.1:LOC100159033	-4,64	0
ACYPI007160	NW_003384246.1:LOC100569674	-1,95	0,01
ACYPI007164	NW_003383543.1:LOC100166281	1,14	0,02
ACYPI007175	Cks1.2	-2,18	0
ACYPI007194	NW_003383844.1:LOC100166311	-1,49	0
ACYPI007272	NW_003383725.1:LOC100166396	-1,18	0,03
ACYPI007276	Putative smc2_b	-1,63	0,01
ACYPI007284	NW_003383662.1:LOC100166409	-1,49	0,03

Continued on next page

Table S2 – continued from previous page

ID	Gene	$\log_2$(fold change)	q-value
ACYPI007294	NW_003383540.1:LOC100166422	1,77	0
ACYPI007312	NW_003383493.1:LOC100166440	-1,42	0,01
ACYPI007338	NW_003383838.1:LOC100166470	1,89	0
ACYPI007430	NW_003384009.1:LOC100166571	-3,11	0
ACYPI007459	NW_003383518.1:LOC100166601	-3,99	0
ACYPI007463	NW_003383931.1:LOC100166605	2,22	0
ACYPI007464	NW_003384893.1:LOC100166606	-3,1	0
ACYPI007467	NW_003383560.1:LOC100160111	-3,6	0
ACYPI007472	NW_003384658.1:LOC100166615	-1,93	0
ACYPI007474	ACYPIP208499	-1,14	0,02
ACYPI007481	ACYPIG475901	2,67	0
ACYPI007493	NW_003384307.1:LOC100166637	-1,14	0,03
ACYPI007519	NW_003383925.1:LOC100166665	-0,99	0,03
ACYPI007530	Serine Protease Inhibitor 2	1,28	0
ACYPI007537	NW_003384296.1:LOC100166684	1,35	0
ACYPI007538	NW_003383846.1:LOC100166685	0,96	0,04
ACYPI007549	NW_003383730.1:LOC100166697	1,09	0,04
ACYPI007554	NW_003383698.1:LOC100166703	1,14	0,04
ACYPI007556	NW_003383900.1:LOC100166705	-1,84	0,03
ACYPI007568	THAUMATIN 6	-1,37	0
ACYPI007570	NW_003383808.1:LOC100166719	1,19	0,01
ACYPI007575	NW_003383817.1:LOC100166724	2,24	0
ACYPI007589	NW_003384615.1:LOC100166739	1,21	0,01
ACYPI007609	P450 related to CYP4 family	4,75	0
ACYPI007641	NW_003383993.1:LOC100166799	-1,27	0,01
ACYPI007644	NW_003384230.1:LOC100166802	-1,28	0
ACYPI007663	Cyclin B	-3,19	0,01
ACYPI007669	NW_003383610.1:LOC100166829	-3,2	0
ACYPI007672	NW_003383596.1:LOC100166832	-0,97	0,04
ACYPI007675	NW_003384383.1:LOC100166835	-1,05	0,03
ACYPI007689	NW_003384538.1:LOC100166850	-1,11	0,01
ACYPI007695	NW_003383593.1:LOC100166856	-1,72	0,01
ACYPI007707	NW_003384689.1:LOC100166868	-3,64	0
ACYPI007713	NW_003384113.1:LOC100166874	1,15	0,03
ACYPI007753	NW_003383560.1:LOC100166916	-4,53	0
ACYPI007771	NW_003383604.1:LOC100166939	1,31	0
ACYPI007783	NW_003383510.1:LOC100166951	2,02	0
ACYPI007787	NW_003383901.1:LOC100166955	-1,26	0,01
ACYPI007833	NW_003384368.1:LOC100167008	1,77	0
ACYPI007844	NW_003383560.1:LOC100167019	-1,02	0,02
ACYPI007931	NW_003383513.1:LOC100167113	3,67	0
ACYPI007941	NW_003384447.1:LOC100167124	-2	0
ACYPI007948	NW_003384283.1:LOC100167131	-1,22	0,02
ACYPI007977	NW_003383797.1:LOC100167163	-1,92	0
ACYPI008012	NW_003384150.1:LOC100167199	1,36	0
ACYPI008017	NW_003388397.1:LOC100167204	-1,62	0,03
ACYPI008022	Lipase12	1,76	0
ACYPI008043	NW_003384289.1:LOC100167232	-1,54	0,01
ACYPI008092	NW_003383760.1:LOC100167283	-1,63	0
ACYPI008110	NW_003384660.1:LOC100167303	1,19	0,02
ACYPI008128	NW_003384034.1:LOC100167324	-2,31	0
ACYPI008137	NW_003384744.1:LOC100167333	-2,68	0
ACYPI008154	NW_003383516.1:LOC100167353	-3,63	0
ACYPI008173	NW_003383603.1:LOC100167374	1,48	0

Continued on next page

Table S2 – continued from previous page

ID	Gene	$\log_2$(fold change)	q-value
ACYPI008179	NW_003383513.1:LOC100167380	-3,54	0,04
ACYPI008192	NW_003384108.1:LOC100167393	1,65	0
ACYPI008196	NW_003383666.1:LOC100167397	-1,04	0,02
ACYPI008213	NW_003383862.1:LOC100167416	1,63	0
ACYPI008260	NW_003384733.1:LOC100167463	-2,01	0,02
ACYPI008282	P450 (partial) related to CYP4a7	4,48	0
ACYPI008289	NW_003384526.1:LOC100167494	2,8	0
ACYPI008309	NW_003384377.1:LOC100167515	1,95	0
ACYPI008354	NW_003383678.1:LOC100167567	-1,92	0
ACYPI008388	NW_003384362.1:LOC100167606	-1,06	0,03
ACYPI008397	NW_003384599.1:LOC100167617	1,42	0
ACYPI008408	NW_003383592.1:LOC100167629	1,14	0,02
ACYPI008410	NW_003383592.1:LOC100167631	1,22	0,01
ACYPI008428	NW_003383647.1:LOC100167651	2,42	0
ACYPI008433	NW_003383514.1:LOC100167656	-1,86	0
ACYPI008456	NW_003384030.1:LOC100167685	-1,97	0
ACYPI008458	NW_003383714.1:LOC100167687	-4,49	0
ACYPI008462	m3AUGepir3s129g74t1_gene	-2,19	0,01
ACYPI008489	NW_003383567.1:LOC100167722	-1,01	0,04
ACYPI008508	NW_003384590.1:LOC100167741	-1,06	0,02
ACYPI008517	NW_003383838.1:LOC100167751	-1,3	0,01
ACYPI008524	CP	1,99	0,01
ACYPI008538	NW_003384243.1:LOC100167772	1,82	0
ACYPI008551	NW_003383642.1:LOC100165748	1,45	0
ACYPI008570	NW_003384437.1:cp23	1,8	0,04
ACYPI008571	Sp2	1,47	0,01
ACYPI008581	NW_003383585.1:LOC100167823	1,26	0,04
ACYPI008595	NW_003383637.1:LOC100167838	-2,53	0
ACYPI008601	NW_003383562.1:LOC100167844	-1,33	0
ACYPI008608	NW_003383493.1:LOC100167851	-2,44	0
ACYPI008624	NW_003384179.1:LOC100167870	-1,07	0,03
ACYPI008631	NW_003384637.1:LOC100167877	-1,31	0,03
ACYPI008643	P450 related to CYP4a7	3,42	0
ACYPI008655	NW_003384282.1:LOC100167904	-1,58	0
ACYPI008656	NW_003384094.1:LOC100167905	-1,57	0,01
ACYPI008667	NW_003383501.1:LOC100167919	1,21	0,03
ACYPI008688	poly (ADP	-2,58	0
ACYPI008695	NW_003383824.1:LOC100160387	-1,64	0,01
ACYPI008696	Lipase01	1,09	0,01
ACYPI008697	NW_003385386.1:LOC100167951	-1,38	0,01
ACYPI008727	NW_003383841.1:LOC100167985	-1,05	0,02
ACYPI008767	NW_003383706.1:LOC100168031	2,06	0
ACYPI008775	Serine Protease Inhibitor 3	-2,33	0
ACYPI008782	ACYPIP880651	0,95	0,04
ACYPI008799	Histone H2A	-inf	0
ACYPI008852	ACYPIP252754	-1	0,03
ACYPI008858	NW_003383543.1:LOC100168132	-1,25	0,02
ACYPI008868	Lipase06	-1,93	0,01
ACYPI008871	NW_003383596.1:LOC100168145	1,65	0,02
ACYPI008897	NW_003383526.1:LOC100168171	-3,51	0
ACYPI008911	NW_003384299.1:Ance-like	-1,25	0,02
ACYPI008914	NW_003383749.1:LOC100168188	-2,86	0
ACYPI008919	NW_003383799.1:LOC100168194	-1,02	0,05
ACYPI008927	NW_003384525.1:LOC100168203	-1,82	0

Continued on next page

Table S2 – continued from previous page

ID	Gene	$\log_2$(fold change)	q-value
ACYPI008945	NW_003383623.1:LOC100168223	1,66	0,02
ACYPI008955	NW_003384654.1:LOC100168235	1,19	0,02
ACYPI008972	NW_003383506.1:LOC100168252	-2,14	0
ACYPI008983	NW_003384587.1:Fib	-2,4	0
ACYPI008984	NW_003384105.1:LOC100168264	1,72	0
ACYPI008990	NW_003384208.1:LOC100168271	1,23	0
ACYPI008994	NW_003383897.1:LOC100168276	-2,57	0,01
ACYPI008998	NW_003383697.1:LOC100168281	-1,32	0
ACYPI009007	NW_003383560.1:LOC100168291	-1,24	0,03
ACYPI009019	NW_003383495.1:LOC100168305	-1,64	0
ACYPI009021	NW_003384097.1:LOC100168309	-1,47	0,01
ACYPI009022	NW_003383515.1:LOC100168310	-2,58	0
ACYPI009024	NW_003384378.1:LOC100569852	-2,05	0,01
ACYPI009027	P450 related to CYP4a7	1,59	0,03
ACYPI009038	NW_003384246.1:LOC100161794	-1,19	0,02
ACYPI009040	NW_003384386.1:LOC100168329	-1,44	0
ACYPI009042	NW_003384247.1:LOC100569954	2,07	0,01
ACYPI009055	NW_003383730.1:LOC100168346	6,2	0
ACYPI009073	m3AUGepi5s237g43t1_gene	1,09	0,04
ACYPI009083	NW_003384096.1:LOC100168377	-1,18	0,01
ACYPI009105	NW_003383714.1:LOC100168401	-3,68	0
ACYPI009109	NW_003383659.1:Ldca	-1,26	0,03
ACYPI009171	NW_003383580.1:LOC100168475	2,18	0,03
ACYPI009176	NW_003383678.1:LOC100168481	-3,52	0
ACYPI009204	NW_003384529.1:LOC100168511	-3,78	0
ACYPI009215	NW_003383496.1:LOC100168523	1,5	0
ACYPI009222	NW_003383498.1:LOC100168531	4,87	0
ACYPI009228	NW_003383902.1:LOC100168537	-2,11	0
ACYPI009264	NW_003383837.1:LOC100168578	-2,82	0
ACYPI009286	NW_003384094.1:LOC100168603	-1,18	0,01
ACYPI009293	NW_003383493.1:LOC100168610	1,36	0
ACYPI009300	NW_003383725.1:LOC100168618	-1,01	0,03
ACYPI009326	NW_003383507.1:LOC100159421	-1,31	0,01
ACYPI009366	ACYPIP802416	-1,87	0,03
ACYPI009378	Polo kinase 1	-3,4	0
ACYPI009388	NW_003384525.1:LOC100168710	-1,45	0,01
ACYPI009391	NW_003383647.1:LOC100168713	1,71	0
ACYPI009394	NW_003383537.1:LOC100571805	1,56	0,01
ACYPI009411	NW_003383561.1:LOC100168734	-1,88	0
ACYPI009456	NW_003384088.1:LOC100168782	3,4	0,04
ACYPI009466	NW_003384225.1:LOC100168792	-2,19	0
ACYPI009480	Oat	-1,32	0
ACYPI009490	NW_003383767.1:LOC100168821	-1,25	0
ACYPI009504	NW_003384138.1:LOC100168835	1,27	0,04
ACYPI009506	NW_003383581.1:LOC100168837	-1,35	0,01
ACYPI009518	NW_003384440.1:LOC100168849	-1,38	0
ACYPI009529	NW_003383666.1:LOC100168861	-1,86	0,01
ACYPI009548	NW_003384175.1:LOC100168884	-1,14	0,01
ACYPI009560	NW_003385285.1:LOC100168897	0,95	0,04
ACYPI009571	NW_003383758.1:IQGAP2	-1,44	0
ACYPI009616	NW_003383598.1:LOC100168953	1,56	0,01
ACYPI009639	NW_003383979.1:Snrpd3	-1,44	0
ACYPI009653	NW_003383527.1:LOC100168995	-3,11	0
ACYPI009656	NW_003384072.1:LOC100168998	-1,7	0,01

Continued on next page

Table S2 – continued from previous page

ID	Gene	$\log_2$(fold change)	q-value
ACYPI009696	NW_003384399.1:LOC100169041	-1,7	0
ACYPI009730	NW_003383813.1:LOC100169079	-1,19	0,02
ACYPI009776	NW_003384447.1:LOC100169129	-1,25	0,03
ACYPI009800	NW_003383626.1:LOC100169156	-1,43	0
ACYPI009820	NW_003383664.1:LOC100169179	-1,2	0,04
ACYPI009849	NW_003383535.1:LOC100169209	-3,87	0
ACYPI009860	NW_003384851.1:LOC100169220	1,09	0,04
ACYPI009919	NW_003384229.1:LOC100169287	1,08	0,04
ACYPI009920	NW_003383593.1:LOC100169288	-2,54	0
ACYPI009926	NW_003383526.1:LOC100169294	1,42	0
ACYPI009960	NW_003383948.1:LOC100169332	2,74	0,03
ACYPI009967	NW_003387522.1:LOC100169340	-3,56	0
ACYPI009984	NW_003383914.1:LOC100169357	-2,81	0,02
ACYPI010031	NW_003383726.1:LOC100169409	-1,14	0,05
ACYPI010033	NW_003383816.1:LOC100169411	-3,7	0,03
ACYPI010055	ACYPIP949557	1,63	0
ACYPI010080	NW_003383567.1:LOC100169465	-1,98	0
ACYPI010088	NW_003384108.1:LOC100169473	1,44	0,03
ACYPI010092	NW_003383678.1:LOC100169477	-1,58	0,01
ACYPI010104	NW_003384699.1:LOC100169489	-1,44	0,01
ACYPI010136	NW_003384733.1:LOC100169525	-0,93	0,04
ACYPI010137	NW_003383516.1:LOC100169526	1,73	0,03
ACYPI010150	NW_003406627.1:LOC100169540	-2,43	0,01
ACYPI010182	NW_003383607.1:LOC100169576	1,34	0,05
ACYPI010185	NW_003383749.1:Rbm34	-1,18	0,03
ACYPI010207	Lipase maturation factor 02	-1,23	0,01
ACYPI010219	NW_003383765.1:LOC100169616	-1,26	0
ACYPI010222	NW_003383908.1:LOC100169619	1,02	0,04
ACYPI010230	NW_003384150.1:LOC100169629	2,58	0
ACYPI010234	NW_003383520.1:LOC100572384	-1,82	0
ACYPI060024	m3AUGpeaaphid0s472g16t1_gene	-1,21	0,03
ACYPI060209	m3AUGepir3s14551g113t1_gene	-1,83	0,01
ACYPI060211	m3AUGepir3s108g3t1_gene	1,98	0,05
ACYPI060476	m3AUGepir3s8457g61t1_gene	1,25	0,03
ACYPI060517	m3AUGepi5s1045g156t1_gene	1,73	0
ACYPI060570	NW_003383671.1:LOC100572798	1,42	0,01
ACYPI060642	NW_003383647.1:LOC100575306	-2,66	0
ACYPI060701	NW_003383583.1:LOC100163851	-1,99	0
ACYPI060710	NW_003383789.1:LOC100163383	-2,68	0
ACYPI060725	NW_003383567.1:LOC100572887	1,49	0,01
ACYPI060728	NW_003383811.1:LOC100575458	-2,47	0,01
ACYPI060811	NW_003383725.1:LOC100168550	1,71	0
ACYPI060813	NW_003383825.1:LOC100166254	-1,21	0,01
ACYPI061426	NW_003383628.1:Mlpt	-4,18	0
ACYPI061436	NW_003383590.1:LOC100571168	1,7	0
ACYPI061444	NW_003383613.1:LOC100302493	-2,03	0,02
ACYPI061445	NW_003387610.1:LOC100169462	-2,71	0
ACYPI061455	NW_003384060.1:LOC100568643	-2,21	0
ACYPI061495	NW_003384978.1:LOC100159353	1,21	0,03
ACYPI061548	NW_003383573.1:LOC100570590	-3,63	0,02
ACYPI061557	NW_003383678.1:LOC100573626	1,53	0,04
ACYPI061569	NW_003383638.1:LOC100571479	1,26	0
ACYPI061574	NW_003384303.1:LOC100570372	-3,08	0
ACYPI061578	NW_003385734.1:LOC100161963	1,37	0,04

Continued on next page

Table S2 – continued from previous page

ID	Gene	$\log_2$(fold change)	q-value
ACYPI061579	NW_003404383.1:LOC100574076	-3,1	0
ACYPI061591	NW_003384850.1:LOC100574375	2,31	0,03
ACYPI061621	NW_003383645.1:LOC100163778	-1,78	0,04
ACYPI061642	NW_003383517.1:LOC100569861	1,71	0,01
ACYPI061660	NW_003383787.1:LOC100571774	inf	0
ACYPI061679	NW_003384849.1:LOC100574218	-2,32	0,03
ACYPI061705	NW_003385090.1:LOC100161475	1,44	0
ACYPI062319	NW_003384553.1:LOC100571626	1,69	0,02
ACYPI062333	NW_003383532.1:LOC100568886	2,72	0
ACYPI062372	NW_003383746.1:LOC100575632	2,18	0,04
ACYPI062389	NW_003383546.1:LOC100570619	-1,63	0,02
ACYPI062434	NW_003383857.1:LOC100575432	-1,61	0
ACYPI062462	NW_003383897.1:LOC100161473	-1,69	0,03
ACYPI062481	NW_003383615.1:LOC100572421	1,19	0,02
ACYPI062512	NW_003385032.1:LOC100574191	1,27	0,01
ACYPI062553	NW_003383660.1:LOC100574362	4,32	0,02
ACYPI062569	NW_003383675.1:LOC100575626	-2,83	0
ACYPI062592	phosphoserine transaminase	-1,75	0
ACYPI062595	ACYPIP222918	1,77	0
ACYPI062743	m3AUGepi4p2s1g6t1_gene	1,22	0,02
ACYPI062831	m3AUGepi5s120g66t1_gene	1,78	0,01
ACYPI063033	m3AUGepir2s653g101t1_gene	-2,91	0
ACYPI063217	NW_003383725.1:LOC100572596	1,68	0
ACYPI063236	NW_003383610.1:Tapdelta	-1,1	0,02
ACYPI063265	NW_003383637.1:LOC100158942	-1,13	0,03
ACYPI063333	NW_003383946.1:LOC100167538	1,85	0,01
ACYPI063340	NW_003385764.1:LOC100168312	-2,42	0,01
ACYPI063370	NW_003383707.1:LOC100164841	1,27	0,04
ACYPI063380	NW_003384055.1:LOC100569548	1,85	0
ACYPI063385	NW_003384333.1:LOC100571146	2,86	0
ACYPI063412	NW_003383667.1:LOC100569691	-1,88	0,02
ACYPI063430	NW_003383631.1:LOC100302371	1,09	0,03
ACYPI063697	m3AUGepi5s117g27t1_gene	1,23	0,02
ACYPI064009	m3AUGepi5s465g21t1_gene	1,79	0
ACYPI064024	NW_003383625.1:LOC100159828	-2,59	0
ACYPI064036	NW_003384457.1:LOC100569699	-2,48	0,02
ACYPI064137	NW_003386721.1:LOC100169214	2,23	0,04
ACYPI064162	NW_003383975.1:LOC100568593	1,61	0,04
ACYPI064275	NW_003385335.1:LOC100573217	2,29	0
ACYPI064286	sp2	1,31	0,02
ACYPI064958	NW_003383590.1:LOC100573995	3,48	0,02
ACYPI064973	NW_003383732.1:LOC100158866	-1,12	0,04
ACYPI064984	NW_003384378.1:LOC100569676	inf	0
ACYPI065022	NW_003383654.1:LOC100571768	1,57	0,02
ACYPI065085	NW_003383717.1:LOC100572514	2,12	0,04
ACYPI065159	NW_003384914.1:LOC100569490	1,36	0,04
ACYPI065196	NW_003405380.1:LOC100569565	-inf	0,01
ACYPI065852	NW_003383660.1:LOC100573655	2,48	0
ACYPI065875	NW_003383653.1:LOC100570649	2,77	0
ACYPI065936	NW_003383730.1:LOC100575537	1,2	0,02
ACYPI065945	NW_003385029.1:LOC100573519	2,21	0,03
ACYPI065960	NW_003383497.1:LOC100576021	-1,23	0,03
ACYPI065989	NW_003383536.1:LOC100161786	2,43	0,04
ACYPI066007	NW_003384064.1:LOC100166912	3,77	0,01

Continued on next page

Table S2 – continued from previous page

ID	Gene	$\log_2$(fold change)	q-value
ACYPI066012	NW_003383659.1:LOC100168716	1,51	0
ACYPI066018	NW_003383567.1:LOC100572954	1,2	0,02
ACYPI066030	NW_003383762.1:LOC100575310	1,26	0,01
ACYPI066036	NW_003383643.1:LOC100574754	-2,02	0,02
ACYPI066073	NW_003383536.1:LOC100575362	-1,63	0,03
ACYPI066091	phantom	4,02	0
ACYPI066680	NW_003383573.1:LOC100575873	-inf	0,02
ACYPI066775	NW_003383697.1:LOC100574621	-3,21	0
ACYPI066784	NW_003384681.1:LOC100161636	-2,13	0,01
ACYPI066829	NW_003384055.1:LOC100569366	1,34	0
ACYPI066831	NW_003383590.1:LOC100571349	1,15	0,04
ACYPI066855	NW_003384156.1:LOC100168774	-1,35	0,01
ACYPI066977	sp3	1,61	0
ACYPI067126	m3AUGepir3s107g58t1_gene	-1,69	0
ACYPI067548	NW_003383598.1:LOC100575510	2,09	0,02
ACYPI067624	NW_003385134.1:LOC100569075	inf	0,02
ACYPI067636	NW_003383887.1:LOC100573424	4,17	0
ACYPI067652	NW_003399439.1:LOC100573648	-1,8	0,01
ACYPI067665	NW_003383492.1:LOC100168392	-1,85	0,01
ACYPI067666	NW_003387415.1:LOC100162128	-2,93	0,03
ACYPI067728	NW_003384337.1:LOC100188938	-1,19	0,01
ACYPI067736	NW_003383675.1:LOC100166877	-1,13	0,05
ACYPI067763	NW_003383714.1:LOC100572989	-1,1	0,02
ACYPI067801	NW_003383751.1:LOC100569356	1,18	0,05
ACYPI067812	NW_003384333.1:LOC100162717	-1,52	0,03
ACYPI067894	m3AUGaphid0s10540g131t1_gene	1,46	0,01
ACYPI068412	NW_003383668.1:LOC100571078	inf	0,04
ACYPI068513	NW_003384369.1:LOC100572525	1,02	0,05
ACYPI068514	NW_003384182.1:LOC100570076	2,41	0
ACYPI068559	NW_003384041.1:LOC100165980	-3,09	0,01
ACYPI068604	NW_003383774.1:LOC100574731	1,75	0,05
ACYPI068652	NW_003383591.1:LOC100160947	-1,88	0,05
ACYPI068670	NW_003384925.1:LOC100570376	1,6	0
ACYPI068863	m3AUGepir3s67g47t1_gene	-1,47	0,01
ACYPI068945	m3AUGaphid0s1064g13t1_gene	1,73	0,03
ACYPI069010	m3AUGepir2s10162g76t1_gene	1,32	0,02
ACYPI069022	m3AUGepir2s89g3t1_gene	2,15	0,01
ACYPI069103	PHUM460950	1,68	0,02
ACYPI069318	NW_003383580.1:LOC100572440	-1,6	0,01
ACYPI069330	NW_003383887.1:LOC100572968	-3,78	0
ACYPI069397	NW_003383590.1:LOC100571007	2,4	0
ACYPI069459	NW_003383887.1:LOC100160747	-2,23	0,01
ACYPI069460	NW_003393919.1:LOC100572108	3,74	0
ACYPI069950	m3AUGepir2s3307g14t1_gene	1,77	0
ACYPI070295	NW_003384021.1:LOC100571241	1,61	0,04
ACYPI070346	NW_003383507.1:LOC100573063	3,24	0,01
ACYPI070362	NW_003383841.1:Bud31	-1,01	0,03
ACYPI070380	NW_003384391.1:LOC100162812	-1,07	0,02
ACYPI070395	NW_003384535.1:LOC100164061	-1,44	0,02
ACYPI070469	NW_003384052.1:LOC100167307	-1,23	0,01
ACYPI070485	NW_003383595.1:LOC100574413	1,69	0
ACYPI071053	m3AUGaphid0s6891g46t1_gene	2,5	0,01
ACYPI071112	NW_003385192.1:LOC100572500	1,31	0,01
ACYPI071130	NW_003384751.1:LOC100569787	1,98	0,03

Continued on next page

Table S2 – continued from previous page

ID	Gene	$\log_2$(fold change)	q-value
ACYPI071192	NW_003385032.1:LOC100574293	2,29	0,01
ACYPI071204	NW_003384893.1:LOC100166780	-4,52	0
ACYPI071217	NW_003383580.1:LOC100161092	-1,28	0,01
ACYPI071231	NW_003383868.1:LOC100302488	-0,99	0,03
ACYPI071378	NW_003383585.1:LOC100570391	-1,65	0,02
ACYPI071954	NW_003383738.1:LOC100169187	-1,92	0
ACYPI072049	NW_003385201.1:LOC100164591	-1,33	0,01
ACYPI072103	NW_003384208.1:LOC100162574	1,26	0,02
ACYPI072138	NW_003384268.1:LOC100573582	1,28	0,01
ACYPI072223	NW_003383660.1:LOC100164308	2,71	0
ACYPI072309	m3AUGepir2s501g112t1_gene	inf	0
ACYPI072356	m3AUGepir2s997g8t1_gene	1,19	0,01
ACYPI072808	NW_003383839.1:LOC100166257	-1,41	0
ACYPI072819	NW_003383756.1:LOC100569783	1,55	0,01
ACYPI072820	NW_003385387.1:LOC100574431	inf	0
ACYPI072835	NW_003385415.1:LOC100167103	-1,64	0,02
ACYPI072839	NW_003383582.1:LOC100570950	1,27	0,03
ACYPI072866	NW_003383596.1:LOC100571729	-1,74	0
ACYPI072868	NW_003384138.1:LOC100575435	1,27	0,02
ACYPI072891	NW_003383522.1:LOC100572756	1,89	0,02
ACYPI072908	NW_003383887.1:LOC100572394	-4,88	0
ACYPI072962	NW_003383548.1:LOC100164556	1,89	0
ACYPI072988	NW_003383648.1:LOC100167045	1,41	0,03
ACYPI073017	NW_003384559.1:LOC100573905	1,19	0,02
ACYPI073243	m3AUGepir2s397g94t1_gene	-3	0
ACYPI073296	m3AUGepi5s8g193t1_gene	-1,65	0,01
ACYPI073422	m3AUGepir3s326g42t1_gene	2,24	0,01
ACYPI073510	m3AUGepi5p2s1g9t1_gene	1,63	0,01
ACYPI073661	NW_003383495.1:LOC100168611	1,03	0,05
ACYPI073713	NW_003383493.1:LOC100162435	-1,34	0
ACYPI073732	NW_003384331.1:LOC100569100	2,18	0
ACYPI073783	NW_003385932.1:LOC100575120	5,17	0
ACYPI073881	NW_003397865.1:LOC100166609	-3,65	0
ACYPI20093	NW_003384451.1:LOC100575038	-1,48	0,02
ACYPI20176	NW_003383657.1:LOC100570627	1,17	0,02
ACYPI20273	NW_003384221.1:LOC100574243	3,26	0,03
ACYPI20430	NW_003385032.1:LOC100573697	1,34	0
ACYPI20476	NW_003384257.1:LOC100569281	-1,73	0
ACYPI20477	m3AUGepir3s768g85t1_gene	-1,75	0
ACYPI20918	NW_003383593.1:LOC100569479	1,01	0,03
ACYPI21412	NW_003383567.1:LOC100572792	1,48	0
ACYPI21413	m3AUGepir3s78g49t1_gene	1,19	0,03
ACYPI21657	NW_003383633.1:LOC100570813	-2,77	0
ACYPI21666	NW_003384080.1:LOC100574710	1,26	0
ACYPI21682	NW_003384392.1:LOC100569955	1,88	0
ACYPI21790	NW_003383518.1:LOC100574437	1,38	0
ACYPI21842	NW_003383504.1:LOC100575100	-1,31	0,05
ACYPI21891	NW_003383878.1:LOC100572124	1,27	0
ACYPI21941	NW_003383554.1:LOC100576027	6,51	0,03
ACYPI21942	m3AUGepi5s65g35t1_gene	inf	0
ACYPI22287	NW_003383618.1:LOC100575400	-2,04	0,01
ACYPI22344	NW_003383792.1:LOC100570601	1,7	0
ACYPI22477	NW_003384085.1:LOC100569913	1,23	0
ACYPI22605	NW_003383577.1:LOC100570553	1,33	0

Continued on next page

Table S2 – continued from previous page

ID	Gene	$\log_2$(fold change)	q-value
ACYPI23019	NW_003384096.1:LOC100570269	-1,71	0,02
ACYPI23163	m3AUGaphid0s166g41t1_gene	1,97	0
ACYPI23734	m3AUGaphid0s230g2t1_gene	2,55	0
ACYPI23752	NW_003383557.1:LOC100574112	1,39	0
ACYPI24218	NW_003383520.1:LOC100568787	2,74	0
ACYPI24233	NW_003383903.1:LOC100570559	0,99	0,04
ACYPI24281	NW_003383908.1:LOC100570990	1,12	0,01
ACYPI24622	NW_003384247.1:LOC100570050	1,52	0
ACYPI24737	NW_003383536.1:LOC100569476	1,93	0,03
ACYPI24817	NW_003383668.1:LOC100571447	1,72	0
ACYPI25151	NW_003383867.1:LOC100571776	1,7	0,02
ACYPI25905	NW_003384344.1:LOC100570824	2,02	0,04
ACYPI25940	NW_003384445.1:LOC100570826	1,38	0,02
ACYPI26018	NW_003383635.1:LOC100570881	1,61	0
ACYPI26315	NW_003384823.1:LOC100575266	-1,95	0,01
ACYPI26368	NW_003384600.1:LOC100575323	-2,54	0
ACYPI26727	NW_003383706.1:LOC100574730	2,88	0
ACYPI26738	NW_003383533.1:LOC100573316	1,3	0
ACYPI26959	NW_003384303.1:LOC100570271	1,8	0
ACYPI27227	NW_003383836.1:LOC100570755	1,02	0,03
ACYPI27259	NW_003383678.1:LOC100575004	2,36	0
ACYPI27808	NW_003383664.1:LOC100568923	-3,12	0
ACYPI27822	NW_003383664.1:LOC100569483	1,68	0
ACYPI27909	NW_003384633.1:LOC100569160	1,5	0
ACYPI28316	NW_003383948.1:LOC100569764	2,62	0
ACYPI28317	NW_003383948.1:LOC100569669	2,32	0,01
ACYPI28869	NW_003384519.1:LOC100159902	1,18	0,01
ACYPI29136	m3AUGepir2s14g112t1_gene	8,1	0
ACYPI29244	NW_003383499.1:LOC100568818	-1,04	0,04
ACYPI29397	m3AUGepi5s213g49t1_gene	1,9	0
ACYPI29519	NW_003384247.1:LOC100570307	-1,81	0,04
ACYPI29674	m3AUGepir3s6g72t1_gene	-1,57	0,01
ACYPI29775	NW_003383505.1:LOC100574166	-1,21	0,02
ACYPI29885	NW_003384085.1:LOC100570184	2,29	0
ACYPI29982	NW_003384601.1:LOC100576041	-1,2	0,04
ACYPI30341	NW_003384593.1:LOC100571399	1,3	0,01
ACYPI30492	m3AUGepir3s1018g37t1_gene	3,25	0
ACYPI30700	NW_003384069.1:LOC100573366	inf	0,01
ACYPI30861	NW_003384462.1:LOC100574017	2,68	0,05
ACYPI30879	NW_003383595.1:LOC100161217	1,35	0,05
ACYPI31122	RabX4	3,39	0,03
ACYPI31575	NW_003384252.1:LOC100573780	1,03	0,03
ACYPI31683	ApisOBP10	1,47	0,05
ACYPI31701	NW_003383749.1:LOC100573101	-1,71	0,01
ACYPI31715	NW_003383820.1:LOC100569360	-3,27	0
ACYPI32264	m3AUGepi4s113g159t1_gene	1,51	0,01
ACYPI32486	NW_003383664.1:LOC100575851	2,44	0
ACYPI32968	NW_003384405.1:LOC100575918	1,05	0,03
ACYPI33027	NW_003384395.1:LOC100571997	-1,58	0,04
ACYPI33244	NW_003383696.1:LOC100161264	1,56	0,02
ACYPI33592	m3AUGepi5p2s23g8t1_gene	-2,49	0
ACYPI33619	NW_003383573.1:LOC100575248	-1,07	0,04
ACYPI33988	NW_003383865.1:LOC100568592	3,42	0
ACYPI34320	NW_003383984.1:LOC100575378	-1,65	0,03

Continued on next page

Table S2 – continued from previous page

ID	Gene	$\log_2$(fold change)	q-value
ACYPI34974	NW_003385230.1:LOC100575574	1,21	0,01
ACYPI34996	NW_003383492.1:LOC100575021	1,46	0,01
ACYPI35003	NW_003384182.1:LOC100569882	3,32	0
ACYPI35004	NW_003385070.1:LOC100571600	4,28	0
ACYPI35095	NW_003383584.1:LOC100574813	1,42	0
ACYPI35350	NW_003383595.1:LOC100573233	1,12	0,03
ACYPI35375	NW_003384574.1:LOC100571398	1,25	0,03
ACYPI35543	NW_003383567.1:LOC100573825	3,03	0
ACYPI35894	NW_003384853.1:LOC100575119	2,36	0
ACYPI36017	NW_003384559.1:LOC100573979	1,25	0,02
ACYPI36033	NW_003384559.1:LOC100573813	1,82	0
ACYPI36528	NW_003384072.1:LOC100570465	-1,69	0
ACYPI36655	NW_003383668.1:LOC100571640	1,59	0
ACYPI36660	NW_003384406.1:LOC100568936	4,13	0,01
ACYPI36730	NW_003383539.1:LOC100569533	2,93	0
ACYPI36734	NW_003383539.1:LOC100569709	0,96	0,04
ACYPI37104	NW_003383666.1:LOC100575193	-2,14	0,01
ACYPI37207	NW_003383726.1:LOC100575108	-1,99	0,02
ACYPI37213	NW_003383726.1:LOC100574673	-1,54	0,01
ACYPI37215	NW_003383726.1:LOC100568588	-3,23	0
ACYPI37616	m3AUGepir2s126g8t1_gene	-6,39	0
ACYPI37959	NW_003383714.1:LOC100575339	2,23	0
ACYPI38156	NW_003383918.1:LOC100572661	-1,45	0
ACYPI38292	NW_003383975.1:LOC100568675	1,93	0
ACYPI38342	NW_003383522.1:LOC100163616	1,87	0
ACYPI38746	NW_003383510.1:LOC100569963	2,15	0
ACYPI38901	NW_003383697.1:LOC100574210	-1,21	0,05
ACYPI38979	NW_003383667.1:LOC100569445	3,66	0
ACYPI38993	NW_003383678.1:LOC100573296	1,11	0,02
ACYPI39140	NW_003383510.1:LOC100574721	2,31	0
ACYPI39361	NW_003383545.1:LOC100573621	-1,48	0,01
ACYPI39369	NW_003383545.1:LOC100573794	1,6	0,03
ACYPI39604	NW_003383787.1:LOC100571580	4,21	0
ACYPI39656	NW_003383904.1:LOC100164169	1,59	0,01
ACYPI40688	NW_003383710.1:LOC100571674	1,18	0,01

Table S3: DAVID clustering tool results table. Clusters include differentially expressed genes (LD vs SD_{14} photoperiod) with similar annotations. The number of genes in each cluster is indicated.

Cluster 1	Enrichment Score: 3.5784187	Genes
	integral component of membrane, Transmembrane helix, Membrane	60
Cluster 2	Enrichment Score: 2.86369156	Genes
	Trypsin- like cysteine/serine peptidase domain	9
	Serine protease	6
	Peptidase S1A, chymotrypsin-type	7
	Peptidase S1, serine-type endopeptidase activity, Tryp_Spc	8
	Protease	8
	Peptidase S1, trypsin family, active site	6
	Hydrolase	13
	Disulfide bond	8
Cluster 3	Enrichment Score: 2.228133402	Genes
	Peptidase C1A, papain C-terminal, Lysosome	5
	cysteine-type peptidase activity	4
Cluster 4	Enrichment Score: 2.01609325	Genes
	Lipid metabolism	7
	Peroxisome	6

	Lipid biosynthesis	6
	Male sterility, NAD-binding, NADP, Fatty acyl-CoA reductase, fatty-acyl-CoA reductase (alcohol-forming) activity	4
	Oxidoreductase	8
	NAD(P)-binding domain	7
Cluster 5	Enrichment Score: 1.70901848	Genes
	transferase activity, transferring acyl groups other than amino-acyl groups	4
	NRF, Acyltransferase 3, Nose resistant-to-fluoxetine protein, N-terminal	3
Cluster 6	Enrichment Score: 1.4175156	Genes
	Glucose-methanol-choline oxidoreductase, N-terminal and C-terminal, acting on CH-OH group of donors, Pyridine nucleotide-disulphide oxidoreductase, FAD/NAD(P)-binding domain, flavin adenine dinucleotide binding	4
Cluster 7	Enrichment Score: 1.04272468	Genes
	Pyridoxal phosphate-dependent transferase, major region, subdomain 1 and 2, pyridoxal phosphate binding	3
Cluster 8	Enrichment Score: 0.9375774	Genes
	Cellular retinaldehyde binding/alpha-tocopherol transport, CRAL/TRIO, N-terminal domain, SEC14	3

	transporter activity	4
	intracellular	4
Cluster 9	Enrichment Score: 0.83747507	Genes
	carbohydrate metabolic process, Glycoside hydrolase, superfamily	4
	Glycoside hydrolase, catalytic domain	3
Cluster 10	Enrichment Score: 0.70387856	Genes
	ABC transporter, conserved site, ABC transporter-like, AAA+ ATPase domain	5
	ATPase activity	3
	P-loop containing nucleoside triphosphate hydrolase	10
	Nucleotide-binding	7
	ATP binding	9
Cluster 11	Enrichment Score: 0.55729046	Genes
	Oxidoreductase	8
	Heme, Iron	3
	heme binding	4
	Metal-binding	6
Cluster 12	Enrichment Score: 0.52989230	Genes
	Major facilitator superfamily domain	6
	Major facilitator superfamily, transmembrane transport	3
Cluster 13	Enrichment Score: 0.3761818	Genes

	kelch-like protein, gigaxonin type	3
	Kelch repeat type 1, BTB/Kelch-associated, SM00875, Galactose oxidase, beta-propeller, BTB/POZ-like, BTB/POZ fold	4
Cluster 14	Enrichment Score: 0.27088487	Genes
	Zinc finger, CCHC-type, ZnF_C2HC	3
	zinc ion binding	6
Cluster 15	Enrichment Score: 0.01682861	Genes
	Zinc finger C2H2-type/integrase DNA-binding domain	3
	metal ion binding	5

Table S4: DAVID clustering tool results table. Clusters include differentially expressed genes (LD vs SD_{10} photoperiod) with similar annotations. The number of genes in each cluster is indicated.

Cluster 1	Enrichment Score: 4.756358728752631	Genes
	Transmembrane, Transmembrane helix, Membrane	189
	integral component of membrane	189
Cluster 2	Enrichment Score: 3.320688209896539	Genes
	Serine protease	11
	IPR001314:Peptidase S1A, chymotrypsin-type	13
	Protease	17
	serine-type endopeptidase activity	17
	Peptidase S1, Trypsin-like cysteine/serine peptidase domain	15
	Tryp_SPc	14
	Peptidase S1, trypsin family, active site	11
	Disulfide bond	26
Cluster 3	Enrichment Score: 2.5668565875817486	Genes
	AAA+ ATPase domain	19
	ABC transporter, conserved site	10
	ABC-2 type transporter	6
	ATPase activity	10
Cluster 4	Enrichment Score: 1.9908314882038798	Genes

	AAA+ ATPase domain	19
	P-loop containing nucleoside triphosphate hydrolase	49
	Nucleotide-binding	38
	ATP-binding	31
	ATP binding	50
Cluster 5	Enrichment Score: 1.8281370192701896	Genes
	Secondary metabolites biosynthesis, transport, and catabolism	12
	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	11
	Cytochrome P450	10
	Cytochrome P450, conserved site, monooxygenase activity, Heme, Iron	9
	iron ion binding	12
	Cytochrome P450, E-class, group I	8
	Oxidoreductase	19
	heme binding	15
Cluster 6	Enrichment Score: 1.7595855126714117	Genes
	DNA replication	10
	DNA replication initiation	5
	MCM complex, Mini-chromosome maintenance, DNA-dependent ATPase	4
	DNA replication	6

	Helicase	8
	Nucleic acid-binding, OB-fold	9
	Mini-chromosome maintenance, conserved site	3
	DNA helicase activity	5
	DNA-binding	10
Cluster 7	Enrichment Score: 1.5005588632762106	Genes
	CRAL/TRIO, N-terminal domain, SEC14, CRAL-TRIO domain	6
	Cellular retinaldehyde binding/alpha-tocopherol transport	5
	transporter activity	9
Cluster 8	Enrichment Score: 1.4613340019340637	Genes
	Serpin family, extracellular space, Serpin domain	5
	Protease inhibitor I4, serpin, conserved site	3
Cluster 9	Enrichment Score: 1.2376433095018415	Genes
	Lipid metabolism	11
	Male sterility, NAD-binding, NADP	6
	Lipid biosynthesis	8
	fatty-acyl-CoA reductase (alcohol-forming) activity	5
	Peroxisome	6
Cluster 10	Enrichment Score: 1.1320663665894866	Genes
	Peptidase C1A, papain C-terminal, Pept_C1, cysteine-type peptidase activity	6

	Lysosome	11
Cluster 11	Enrichment Score: 1.026038134135763	Genes
	pyridoxal phosphate binding	6
	Pyridoxal phosphate-dependent transferase, major region, subdomain 1	5
	Pyridoxal phosphate-dependent transferase, major region, subdomain 2	4
Cluster 12	Enrichment Score: 0.8977779943909802	Genes
	C-type lectin-like, C-type lectin fold	3
Cluster 13	Enrichment Score: 0.8671897672419997	Genes
	Kinesin, motor region, conserved site, KISc	4
	microtubule motor activity	6
	microtubule-based movement	5
	Microtubule	5
	Motor protein	3
Cluster 14	Enrichment Score: 0.8152608168040392	Genes
	Adenylyl cyclase class-3/4/guanylyl cyclase, conserved site	3
	Lyase	5
	intracellular signal transduction	5
Cluster 15	Enrichment Score: 0.7869649949757057	Genes
	Calcium	5

	Cadherin, homophilic cell adhesion via plasma membrane adhesion molecules	3
	plasma membrane	6
Cluster 16	Enrichment Score: 0.7371516635130186	Genes
	Glucose-methanol-choline oxidoreductase, N-terminal and C-terminal, acting on CH-OH group of donors, Pyridine nucleotide-disulphide oxidoreductase, FAD/NAD(P)-binding domain	5
	Glucose-methanol-choline oxidoreductase, C-terminal	5
	flavin adenine dinucleotide binding	6
Cluster 17	Enrichment Score: 0.7028761804157181	Genes
	RNA helicase, DEAD-box type, Q motif	5
	Helicase, C-terminal, superfamily 1/2, ATP-binding domain, DEXDc	9
	DNA/RNA helicase, DEAD/DEAH box type, N-terminal	7
	HELICc	8
	helicase activity	4
	RNA helicase, ATP-dependent, DEAD-box, conserved site	3
Cluster 18	Enrichment Score: 0.621411657926761	Genes
	WD40/YVTN repeat-like-containing domain	17
	WD40 repeat, conserved site	9

	WD40	14
	WD40-repeat-containing domain	15
	G-protein beta WD-40 repeat	5
Cluster 19	Enrichment Score: 0.6059935989510573	Genes
	Glutathione metabolism	8
	Peptidase M1, membrane alanine aminopeptidase, N-terminal, aminopeptidase/leukotriene A4 hydrolase, DUF3358	4
	metallopeptidase activity	5
Cluster 20	Enrichment Score: 0.5961784293983855	Genes
	Lipid biosynthesis	8
	Fatty acid biosynthesis/metabolism, 3-oxo-cerotoyl-CoA, 3-oxo-lignoceronyl-CoA and 3-oxo-arachidoyl-CoA synthase activity, GNS1/SUR4 membrane protein	3
Cluster 21	Enrichment Score: 0.5104673795076791	Genes
	GPCR, family 2-like	5
	GPCR, family 2, secretin-like, G-protein coupled receptor activity	3
	cell surface receptor signaling pathway	4
	Receptor	7
Cluster 22	Enrichment Score: 0.4686573469102974	Genes
	General substrate transporter	7
	transmembrane transporter activity	5

	Sugar transporter, conserved site	4
Cluster 23	Enrichment Score: 0.46372311167305835	Genes
	Acetyl-coenzyme A transporter 1, acetyl-CoA transporter activity	3
	Glycosphingolipid biosynthesis - ganglio series	4
Cluster 24	Enrichment Score: 0.451820096492242	Genes
	Aldolase-type TIM barrel	6
	Glycosyl hydrolase, family 13, catalytic domain, Aamy, Glycoside hydrolase, catalytic domain	3
	Glycoside hydrolase, superfamily	6
	carbohydrate metabolic process	9
	catalytic activity	9
Cluster 25	Enrichment Score: 0.4422297243354925	Genes
	Ribonucleoprotein LSM domain	3
Cluster 26	Enrichment Score: 0.42175955816650323	Genes
	translation initiation factor activity, Protein biosynthesis, PINT, eukaryotic translation initiation factor 3 complex, eukaryotic 43S and 48s preinitiation complex, Proteasome component (PCI) domain, regulation of translational initiation, formation of translation preinitiation complex	3
	Cytoplasm	8
Cluster 27	Enrichment Score: 0.405254634183642	Genes

	Peptidase C19, ubiquitin carboxyl-terminal hydrolase 2, protein deubiquitination	4
	Peptidase C19, ubiquitin carboxyl-terminal hydrolase 2, conserved site, ubiquitin-dependent protein catabolic process	3
Cluster 28	Enrichment Score: 0.39166253872466O4	Genes
	Metal-binding	32
	Zinc-finger	14
	Zinc	18
Cluster 29	Enrichment Score: 0.34355409368416273	Genes
	Kelch-like protein, gigaxonin	7
	Kelch repeat type 1, Galactose oxidase, beta-propeller	10
	BTB/POZ-like, BTB/POZ fold	12
	BTB/Kelch-associated, SM00875	10
Cluster 30	Enrichment Score: 0.32626141814280S5	Genes
	Domain of unknown function DUF4371	12
	Zinc finger, TTF-type	4
Cluster 31	Enrichment Score: 0.30800482991998457	Genes
	ChtBD2, Chitin binding domain, chitin binding, chitin metabolic process	3
	extracellular region	6
Cluster 32	Enrichment Score: 0.29379853292909O5	Genes
	calcium ion binding	12

	EF-Hand 1, calcium-binding site	4
	EF-hand-like domain	5
Cluster 33	Enrichment Score: 0.2828053116974681	Genes
	Leucine-rich repeat, typical subtype	6
Cluster 34	Enrichment Score: 0.2625060586015743	Genes
	0016758∼transferase activity, transferring hexosyl groups	4
	UDP-glucuronosyl/UDP-glucosyltransferase	5
	metabolic process	10
	Ascorbate and aldarate metabolism	5
	Porphyrin and chlorophyll metabolism, Pentose and glucuronate interconversions, Drug metabolism - other enzymes	5
	Drug metabolism - cytochrome P450, Metabolism of xenobiotics by cytochrome P450, Retinol metabolism	4
Cluster 35	Enrichment Score: 0.2538688236158045	Genes
	response to oxidative stress	5
	Haem peroxidase, animal, subgroup	4
Cluster 36	Enrichment Score: 0.2520595198784673	Genes
	Dbl homology (DH) domain, Rho guanyl-nucleotide exchange factor activity, regulation of Rho protein signal transduction, SM00233:PH,	3
	Pleckstrin homology domain	4

	Pleckstrin homology-like domain	7
Cluster 37	Enrichment Score: 0.2375752840908521	Genes
	Insulin-like growth factor binding protein, N-terminal	3
	Epidermal growth factor-like domain, EGF	4
Cluster 38	Enrichment Score: 0.22306455288223578	Genes
	small GTPase mediated signal transduction	7
	Small GTPase superfamily	5
	Small GTP-binding protein domain	6
	GTP binding	9
Cluster 39	Enrichment Score: 0.21915465652000046	Genes
	Histone core, Nucleosome core, Histone-fold	3
	Chromosome	4
Cluster 40	Enrichment Score: 0.15704547406259642	Genes
	Low-density lipoprotein (LDL) receptor class A repeat, LDLa, (LDL) receptor class A, conserved site	3
Cluster 41	Enrichment Score: 0.10357333251387933	Genes
	Ankyrin repeat	4
	Ankyrin repeat-containing domain	6
Cluster 42	Enrichment Score: 0.08721130079354159	Genes
	Carbon metabolism	6
	Biosynthesis of antibiotics	10
	Biosynthesis of amino acids	4
Cluster 43	Enrichment Score: 0.07585537478010773	Genes

	protein serine/threonine kinase activity	5
	Serine/threonine-protein kinase, active site	7
	S_TKc	8
	protein kinase activity	4
	Protein kinase, catalytic domain	9
	Kinase	6
	Protein kinase, ATP binding site	6
	Protein kinase-like domain	10
Cluster 44	Enrichment Score: 0.0701822011104643	Genes
	RRM, RNA recognition motif domain, Nucleotide-binding, alpha-beta plait	7
	nucleotide binding	9
Cluster 45	Enrichment Score: 0.04233206317310632	Genes
	Transcription	5
	transcription, DNA-templated	8
	transcription factor activity, sequence-specific DNA binding	6
	Transcription regulation	3
	sequence-specific DNA binding	4
Cluster 46	Enrichment Score: 0.01977271952182817	Genes
	Transposase protein, transposase activity, transposition, DNA-mediated	3
Cluster 47	Enrichment Score: 0.004326885259436021	Genes

	RING	3
	Zinc finger, RING/FYVE/PHD-type	6
Cluster 48	Enrichment Score: 8.254752317507158E-4	Genes
	metal ion binding	20
	Zinc finger, C2H2	10
	Zinc finger, C2H2-like, C2H2-type/integrase DNA-binding domain	8

Table S5: REVIGO results for enriched GO terms when comparing LD vs SD_{14} photoperiods. See the text for the definition of frequency (Freq.) and uniqueness (Uniq.).

GO term ID	Description	Freq.	Uniq.
GO:0006200	(obsolete) ATP catabolic process	0.143%	0,985
GO:0006898	receptor-mediated endocytosis	0.095%	0,931
GO:0006911	phagocytosis, engulfment	0.013%	0,809
GO:0007108	(obsolete) cytokinesis, initiation of separation	0.143%	0,985
GO:0007160	cell-matrix adhesion	0.051%	0,965
GO:0007631	feeding behavior	0.026%	0,919
GO:0007632	visual behavior	0.012%	0,616
GO:0007617	mating behavior	0.012%	0,674
GO:0008049	male courtship behavior	0.004%	0,688
GO:0008344	adult locomotory behavior	0.024%	0,677
GO:0008152	metabolic process	75.387%	0,996
GO:0009987	cellular process	63.780%	0,994
GO:0030421	defecation	0.002%	0,776
GO:0032501	multicellular organismal process	2.373%	0,985
GO:0035286	(obsolete) leg segmentation	0.143%	0,985
GO:0040007	growth	0.317%	0,985
GO:0040011	locomotion	0.997%	0,985
GO:0042744	hydrogen peroxide catabolic process	0.093%	0,94
GO:0046685	response to arsenic-containing substance	0.020%	0,886
GO:0048066	developmental pigmentation	0.020%	0,907
GO:0008057	eye pigment granule organization	0.001%	0,824
GO:0048511	rhythmic process	0.077%	0,985
GO:0050790	regulation of catalytic activity	1.575%	0,932
GO:0065007	biological regulation	20.498%	0,988
GO:0006030	chitin metabolic process	0.077%	0,932
GO:0006334	nucleosome assembly	0.089%	0,862
GO:0006629	lipid metabolic process	3.522%	0,881
GO:0000281	mitotic cytokinesis	0.070%	0,812
GO:0007110	meiosis I cytokinesis	0.001%	0,758
GO:0007112	male meiosis cytokinesis	0.002%	0,747
GO:0007111	meiosis II cytokinesis	0.001%	0,759
GO:0000915	actomyosin contractile ring assembly	0.015%	0,778
GO:0007568	aging	0.088%	0,747
GO:0005975	carbohydrate metabolic process	5.260%	0,961
GO:0006468	protein phosphorylation	4.137%	0,899
GO:0019483	beta-alanine biosynthetic process	0.000%	0,87
GO:0040039	inductive cell migration	0.001%	0,867

Continued on next page

Table S5 – continued from previous page

GO term ID	Description	Freq.	Uniq.
GO:0042416	dopamine biosynthetic process	0.003%	0,906
GO:0042053	regulation of dopamine metabolic process	0.003%	0,878
GO:0007589	body fluid secretion	0.013%	0,848
GO:0007595	lactation	0.004%	0,647
GO:0048009	insulin-like growth factor receptor signaling pathway	0.007%	0,838
GO:0070588	calcium ion transmembrane transport	0.157%	0,93
GO:0009617	response to bacterium	0.145%	0,879
GO:0010941	regulation of cell death	0.344%	0,841
GO:0010212	response to ionizing radiation	0.040%	0,869
GO:0009416	response to light stimulus	0.157%	0,86
GO:0006591	ornithine metabolic process	0.088%	0,828
GO:0016070	RNA metabolic process	15.951%	0,894
GO:0019722	calcium-mediated signaling	0.040%	0,823
GO:0055114	oxidation-reduction process	15.060%	0,88
GO:0006810	transport	17.616%	0,917
GO:0007479	leg disc proximal/distal pattern formation	0.001%	0,716
GO:0010587	miRNA catabolic process	0.002%	0,916
GO:0006950	response to stress	4.575%	0,878
GO:0006413	translational initiation	0.518%	0,88
GO:0032507	maintenance of protein location in cell	0.057%	0,808
GO:0040003	chitin-based cuticle development	0.004%	0,711
GO:0048067	cuticle pigmentation	0.003%	0,699
GO:0006422	aspartyl-tRNA aminoacylation	0.045%	0,805
GO:0008272	sulfate transport	0.150%	0,862
GO:0006212	uracil catabolic process	0.005%	0,844
GO:0016319	mushroom body development	0.004%	0,694
GO:0045727	positive regulation of translation	0.082%	0,846
GO:0045463	R8 cell development	0.000%	0,683
GO:0016318	ommatidial rotation	0.002%	0,674
GO:0001737	establishment of imaginal disc-derived wing hair orientation	0.001%	0,631
GO:0045467	R7 cell development	0.001%	0,666
GO:0006979	response to oxidative stress	0.575%	0,886
GO:0006508	proteolysis	5.223%	0,931
GO:0008340	determination of adult lifespan	0.020%	0,697
GO:0001708	cell fate specification	0.028%	0,713
GO:0008643	carbohydrate transport	0.682%	0,854
GO:0002119	nematode larval development	0.015%	0,695
GO:0006551	leucine metabolic process	0.162%	0,822

Continued on next page

Table S5 – continued from previous page

GO term ID	Description	Freq.	Uniq.
GO:0007291	sperm individualization	0.003%	0,629
GO:0009072	aromatic amino acid family metabolic process	0.719%	0,801
GO:2000274	regulation of epithelial cell migration, open tracheal system	0.001%	0,669
GO:0007412	axon target recognition	0.002%	0,637
GO:0007552	metamorphosis	0.031%	0,69
GO:0010033	response to organic substance	0.900%	0,859
GO:0045686	negative regulation of glial cell differentiation	0.004%	0,636
GO:0045666	positive regulation of neuron differentiation	0.057%	0,591
GO:0050774	negative regulation of dendrite morphogenesis	0.003%	0,609
GO:0006414	translational elongation	0.777%	0,877
GO:0007605	sensory perception of sound	0.039%	0,741
GO:0050974	detection of mechanical stimulus involved in sensory perception	0.007%	0,669
GO:0048863	stem cell differentiation	0.047%	0,706
GO:0000289	nuclear-transcribed mRNA poly(A) tail shortening	0.018%	0,906
GO:0009792	embryo development ending in birth or egg hatching	0.155%	0,662
GO:0006520	cellular amino acid metabolic process	5.591%	0,778
GO:0014045	establishment of endothelial blood-brain barrier	0.001%	0,68
GO:0008052	sensory organ boundary specification	0.001%	0,705
GO:0006355	regulation of transcription, DNA-templated	9.917%	0,833
GO:0016339	calcium-dependent cell-cell adhesion via plasma membrane cell adhesion molecules	0.004%	0,963
GO:0007156	homophilic cell adhesion via plasma membrane adhesion molecules	0.095%	0,961
GO:0008407	chaeta morphogenesis	0.002%	0,703
GO:0006066	alcohol metabolic process	0.422%	0,87
GO:0016246	RNA interference	0.011%	0,829
GO:0031054	pre-miRNA processing	0.004%	0,721
GO:0022600	digestive system process	0.019%	0,714
GO:0006855	drug transmembrane transport	0.189%	0,765
GO:0016310	phosphorylation	7.764%	0,936
GO:0000022	mitotic spindle elongation	0.006%	0,802

Continued on next page

Table S5 – continued from previous page

GO term ID	Description	Freq.	Uniq.
GO:0051225	spindle assembly	0.046%	0,777
GO:0000212	meiotic spindle organization	0.007%	0,738
GO:0014070	response to organic cyclic compound	0.227%	0,863
GO:0048545	response to steroid hormone	0.118%	0,865
GO:0030036	actin cytoskeleton organization	0.381%	0,771
GO:0007015	actin filament organization	0.240%	0,777
GO:0042594	response to starvation	0.109%	0,874
GO:0007413	axonal fasciculation	0.006%	0,624
GO:0031290	retinal ganglion cell axon guidance	0.006%	0,545
GO:0048841	regulation of axon extension involved in axon guidance	0.004%	0,518
GO:0048675	axon extension	0.025%	0,595
GO:0071356	cellular response to tumor necrosis factor	0.029%	0,857
GO:0001666	response to hypoxia	0.049%	0,864
GO:0071454	cellular response to anoxia	0.002%	0,844
GO:0016476	regulation of embryonic cell shape	0.002%	0,645
GO:0007126	meiotic nuclear division	0.126%	0,712
GO:0040035	hermaphrodite genitalia development	0.001%	0,68
GO:0007281	germ cell development	0.093%	0,577
GO:0008406	gonad development	0.043%	0,621
GO:0048477	oogenesis	0.054%	0,588
GO:0000398	mRNA splicing, via spliceosome	0.315%	0,901
GO:0006865	amino acid transport	0.813%	0,846
GO:0055085	transmembrane transport	8.916%	0,92
GO:0002121	inter-male aggressive behavior	0.004%	0,901
GO:0007455	eye-antennal disc morphogenesis	0.002%	0,696
GO:0007638	mechanosensory behavior	0.004%	0,776

Table S6: REVIGO results for enriched GO terms when comparing LD vs SD_{10} photoperiods. See the text for the definition of frequency (Freq.) and uniqueness (Uniq.).

GO term ID	Description	Freq.	Uniq.
GO:0000090	mitotic anaphase	0.004%	0,952
GO:0051329	mitotic interphase	0.004%	0,945
GO:0000084	mitotic S phase	0.004%	0,952
GO:0000080	mitotic G1 phase	0.004%	0,952
GO:0000085	mitotic G2 phase	0.004%	0,952
GO:0022403	cell cycle phase	0.001%	0,952
GO:0000239	pachytene	0.001%	0,948
GO:0002115	store-operated calcium entry	0.007%	0,962
GO:0007622	rhythmic behavior	0.012%	0,947
GO:0045475	locomotor rhythm	0.006%	0,946
GO:0007617	mating behavior	0.012%	0,569
GO:0001541	ovarian follicle development	0.011%	0,499
GO:0042745	circadian sleep/wake cycle	0.005%	0,947
GO:0016339	calcium-dependent cell-cell adhesion via plasma membrane cell adhesion molecules	0.004%	0,993
GO:0033076	isoquinoline alkaloid metabolic process	0.000%	0,96
GO:0046684	response to pyrethroid	0.000%	0,944
GO:0050916	sensory perception of sweet taste	0.003%	0,665
GO:0060250	germ-line stem-cell niche homeostasis	0.001%	0,918
GO:0051290	protein heterotetramerization	0.004%	0,891
GO:0052314	phytoalexin metabolic process	0.001%	0,897
GO:0007525	somatic muscle development	0.005%	0,651
GO:0008334	histone mRNA metabolic process	0.005%	0,946
GO:0048102	autophagic cell death	0.006%	0,88
GO:0008356	asymmetric cell division	0.012%	0,87
GO:0034199	activation of protein kinase A activity	0.000%	0,892
GO:0010216	maintenance of DNA methylation	0.002%	0,943
GO:0006584	catecholamine metabolic process	0.010%	0,944
GO:0042416	dopamine biosynthetic process	0.003%	0,941
GO:0042427	serotonin biosynthetic process	0.002%	0,943
GO:0040039	inductive cell migration	0.001%	0,861
GO:0019483	beta-alanine biosynthetic process	0.000%	0,888
GO:0008105	asymmetric protein localization	0.008%	0,959
GO:0033979	box H/ACA snoRNA metabolic process	0.002%	0,945
GO:0002385	mucosal immune response	0.004%	0,927
GO:0006384	transcription initiation from RNA polymerase III promoter	0.007%	0,942

Continued on next page

Table S6 – continued from previous page

GO term ID	Description	Freq.	Uniq.
GO:0050428	3'-phosphoadenosine 5'-phosphosulfate biosynthetic process	0.000%	0,875
GO:0010866	regulation of triglyceride biosynthetic process	0.003%	0,857
GO:0008101	decapentaplegic signaling pathway	0.002%	0,856
GO:0006533	aspartate catabolic process	0.001%	0,875
GO:0034769	basement membrane disassembly	0.001%	0,847
GO:0015782	CMP-N-acetylneuraminate transport	0.000%	0,909
GO:0015785	UDP-galactose transport	0.003%	0,893
GO:0015789	UDP-N-acetylgalactosamine transport	0.000%	0,902
GO:0007000	nucleolus organization	0.004%	0,889
GO:0006999	nuclear pore organization	0.009%	0,878
GO:0006277	DNA amplification	0.002%	0,938
GO:0009270	response to humidity	0.000%	0,928
GO:0006297	nucleotide-excision repair, gap filling	0.003%	0,898
GO:0060305	regulation of cell diameter	0.001%	0,81
GO:0008057	eye pigment granule organization	0.001%	0,838
GO:0007376	cephalic furrow formation	0.000%	0,618
GO:0007370	ventral furrow formation	0.001%	0,584
GO:0007378	amnioserosa formation	0.000%	0,616
GO:0007377	germ-band extension	0.001%	0,601
GO:0007375	anterior midgut invagination	0.000%	0,599
GO:0016486	peptide hormone processing	0.008%	0,884
GO:0048785	hatching gland development	0.000%	0,616
GO:0031120	snRNA pseudouridine synthesis	0.004%	0,935
GO:0000461	endonucleolytic cleavage to generate mature 3'-end of SSU-rRNA from (SSU-rRNA, 5.8S rRNA, LSU-rRNA)	0.002%	0,884
GO:0035074	pupation	0.000%	0,608
GO:0097352	autophagosome maturation	0.008%	0,82
GO:0061320	pericardial nephrocyte differentiation	0.001%	0,596
GO:0035001	dorsal trunk growth, open tracheal system	0.000%	0,59
GO:0035002	liquid clearance, open tracheal system	0.001%	0,571
GO:0007428	primary branching, open tracheal system	0.000%	0,554
GO:0007427	epithelial cell migration, open tracheal system	0.002%	0,542
GO:0035159	regulation of tube length, open tracheal system	0.002%	0,544

Continued on next page

Table S6 – continued from previous page

GO term ID	Description	Freq.	Uniq.
GO:0035157	negative regulation of fusion cell fate specification	0.000%	0,536
GO:0035158	regulation of tube diameter, open tracheal system	0.001%	0,565
GO:0035152	regulation of tube architecture, open tracheal system	0.005%	0,534
GO:0035151	regulation of tube size, open tracheal system	0.003%	0,538
GO:0035156	fusion cell fate specification	0.000%	0,566
GO:0035155	negative regulation of terminal cell fate specification, open tracheal system	0.000%	0,536
GO:0035147	branch fusion, open tracheal system	0.001%	0,538
GO:0015734	taurine transport	0.006%	0,896
GO:0040034	regulation of development, heterochronic	0.007%	0,645
GO:0045900	negative regulation of translational elongation	0.004%	0,859
GO:0034969	histone arginine methylation	0.003%	0,874
GO:0015014	heparan sulfate proteoglycan biosynthetic process, polysaccharide chain biosynthetic process	0.001%	0,872
GO:0030206	chondroitin sulfate biosynthetic process	0.003%	0,863
GO:0007398	ectoderm development	0.005%	0,639
GO:0007591	molting cycle, chitin-based cuticle	0.005%	0,625
GO:0035188	hatching	0.001%	0,624
GO:0015842	aminergic neurotransmitter loading into synaptic vesicle	0.001%	0,895
GO:0070328	triglyceride homeostasis	0.008%	0,913
GO:0007458	progression of morphogenetic furrow involved in compound eye morphogenesis	0.000%	0,581
GO:0016318	ommatidial rotation	0.002%	0,5
GO:0016330	second mitotic wave involved in compound eye morphogenesis	0.001%	0,527
GO:0045500	sevenless signaling pathway	0.000%	0,512
GO:0046667	compound eye retinal cell programmed cell death	0.001%	0,526
GO:0035071	salivary gland cell autophagic cell death	0.005%	0,495
GO:0007464	R3/R4 cell fate commitment	0.001%	0,514
GO:0046069	cGMP catabolic process	0.001%	0,862
GO:0006198	cAMP catabolic process	0.007%	0,851
GO:0000060	protein import into nucleus, translocation	0.012%	0,891

Continued on next page

Table S6 – continued from previous page

GO term ID	Description	Freq.	Uniq.
GO:0030718	germ-line stem cell population maintenance	0.005%	0,609
GO:0042049	cellular acyl-CoA homeostasis	0.000%	0,873
GO:0045751	negative regulation of Toll signaling pathway	0.001%	0,83
GO:0060472	positive regulation of cortical granule exocytosis by positive regulation of cytosolic calcium ion concentration	0.000%	0,741
GO:0015872	dopamine transport	0.007%	0,896
GO:0045740	positive regulation of DNA replication	0.012%	0,87
GO:0030431	sleep	0.012%	0,612
GO:0006212	uracil catabolic process	0.005%	0,861
GO:0007419	ventral cord development	0.004%	0,57
GO:0007366	periodic partitioning by pair rule gene	0.000%	0,587
GO:0007354	zygotic determination of anterior/posterior axis, embryo	0.002%	0,551
GO:0007367	segment polarity determination	0.002%	0,556
GO:0007365	periodic partitioning	0.002%	0,554
GO:0007314	oocyte anterior/posterior axis specification	0.005%	0,455
GO:0008595	anterior/posterior axis specification, embryo	0.009%	0,526
GO:0035290	trunk segmentation	0.001%	0,56
GO:0035289	posterior head segmentation	0.000%	0,578
GO:0071454	cellular response to anoxia	0.002%	0,905
GO:0042335	cuticle development	0.012%	0,584
GO:0051299	centrosome separation	0.006%	0,79
GO:0007144	female meiosis I	0.002%	0,738
GO:0016321	female meiosis chromosome segregation	0.003%	0,735
GO:0048146	positive regulation of fibroblast proliferation	0.009%	0,866
GO:0033574	response to testosterone	0.003%	0,925
GO:0042688	crystal cell differentiation	0.001%	0,632
GO:0035204	negative regulation of lamellocyte differentiation	0.001%	0,509
GO:0007516	hemocyte development	0.000%	0,625
GO:0035168	larval lymph gland hemocyte differentiation	0.001%	0,54
GO:0051439	regulation of ubiquitin-protein ligase activity involved in mitotic cell cycle	0.001%	0,799
GO:0035217	labial disc development	0.000%	0,571
GO:0007499	ectoderm and mesoderm interaction	0.000%	0,632

Continued on next page

Table S6 – continued from previous page

GO term ID	Description	Freq.	Uniq.
GO:0040035	hermaphrodite genitalia development	0.001%	0,56
GO:0000916	actomyosin contractile ring contraction	0.012%	0,818
GO:0030970	retrograde protein transport, ER to cytosol	0.011%	0,771
GO:0042023	DNA endoreduplication	0.007%	0,809
GO:0006267	pre-replicative complex assembly involved in nuclear cell cycle DNA replication	0.009%	0,763
GO:0032201	telomere maintenance via semi-conservative replication	0.001%	0,753
GO:0007412	axon target recognition	0.002%	0,523
GO:0007089	traversing start control point of cell cycle	0.003%	0,794
GO:0010971	positive regulation of G2/M transition of mitotic cell cycle	0.009%	0,785
GO:0007095	mitotic G2 DNA damage checkpoint	0.012%	0,733
GO:0006977	DNA damage response, signal transduction by p53 class mediator resulting in cell cycle arrest	0.002%	0,736
GO:0043928	exonucleolytic nuclear-transcribed mRNA catabolic process involved in deadenylation-dependent decay	0.006%	0,931
GO:0071377	cellular response to glucagon stimulus	0.002%	0,91
GO:0031018	endocrine pancreas development	0.012%	0,559
GO:0046331	lateral inhibition	0.012%	0,575
GO:0016246	RNA interference	0.011%	0,808
GO:0031054	pre-miRNA processing	0.004%	0,737
GO:0030702	chromatin silencing at centromere	0.009%	0,752
GO:0007285	primary spermatocyte growth	0.000%	0,572
GO:0010002	cardioblast differentiation	0.005%	0,546
GO:0045705	negative regulation of salivary gland boundary specification	0.000%	0,542
GO:0008052	sensory organ boundary specification	0.001%	0,535
GO:0016360	sensory organ precursor cell fate determination	0.001%	0,507
GO:0001700	embryonic development via the syncytial blastoderm	0.009%	0,569
GO:0001964	startle response	0.006%	0,627
GO:0007348	regulation of syncytial blastoderm mitotic cell cycle	0.000%	0,545
GO:0090231	regulation of spindle checkpoint	0.009%	0,779

Continued on next page

Table S6 – continued from previous page

GO term ID	Description	Freq.	Uniq.
GO:0043653	mitochondrial fragmentation involved in apoptotic process	0.002%	0,826
GO:0019732	antifungal humoral response	0.001%	0,914
GO:0010587	miRNA catabolic process	0.002%	0,931
GO:0007494	midgut development	0.003%	0,567
GO:0007442	hindgut morphogenesis	0.004%	0,553
GO:0007443	Malpighian tubule morphogenesis	0.002%	0,524
GO:0007440	foregut morphogenesis	0.003%	0,562
GO:0008347	glial cell migration	0.010%	0,522
GO:0045686	negative regulation of glial cell differentiation	0.004%	0,502
GO:0035269	protein O-linked mannosylation	0.011%	0,854
GO:0045746	negative regulation of Notch signaling pathway	0.009%	0,809
GO:0008302	female germline ring canal formation, actin assembly	0.000%	0,557
GO:0030720	oocyte localization involved in germarium-derived egg chamber formation	0.001%	0,54
GO:0030708	germarium-derived female germ-line cyst encapsulation	0.000%	0,523
GO:0035017	cuticle pattern formation	0.002%	0,579
GO:0031118	rRNA pseudouridine synthesis	0.005%	0,874
GO:0048009	insulin-like growth factor receptor signaling pathway	0.007%	0,847
GO:0006616	SRP-dependent cotranslational protein targeting to membrane, translocation	0.008%	0,801
GO:0007296	vitellogenesis	0.001%	0,567
GO:0030713	ovarian follicle cell stalk formation	0.001%	0,494
GO:0007298	border follicle cell migration	0.007%	0,463
GO:0007307	eggshell chorion gene amplification	0.001%	0,483
GO:0007313	maternal specification of dorsal/ventral axis, oocyte, soma encoded	0.001%	0,511
GO:0046843	dorsal appendage formation	0.003%	0,472
GO:0001556	oocyte maturation	0.006%	0,524
GO:0022416	chaeta development	0.005%	0,56
GO:0040018	positive regulation of multicellular organism growth	0.012%	0,547
GO:0045570	regulation of imaginal disc growth	0.002%	0,51
GO:0048542	lymph gland development	0.003%	0,568
GO:0000022	mitotic spindle elongation	0.006%	0,783
GO:0000212	meiotic spindle organization	0.007%	0,724

Continued on next page

Table S6 – continued from previous page

GO term ID	Description	Freq.	Uniq.
GO:0007451	dorsal/ventral lineage restriction, imaginal disc	0.000%	0,554
GO:0048100	wing disc anterior/posterior pattern formation	0.000%	0,542
GO:0035224	genital disc anterior/posterior pattern formation	0.000%	0,536
GO:0007479	leg disc proximal/distal pattern formation	0.001%	0,529
GO:0007473	disc proximal/distal pattern formation	0.000%	0,548
GO:0048190	wing disc dorsal/ventral pattern formation	0.003%	0,501
GO:0007450	dorsal/ventral pattern formation imaginal disc	0.003%	0,501
GO:0000722	telomere maintenance via recombination	0.011%	0,803
GO:0030241	muscle myosin thick filament assembly	0.001%	0,56
GO:0050774	regulation of dendrite morphogenesis	0.003%	0,481
GO:0007527	adult somatic muscle development	0.001%	0,675
GO:0060810	intracellular mRNA localization involved in pattern specification process	0.003%	0,548
GO:0032355	response to estradiol	0.009%	0,921
GO:0016319	mushroom body development	0.004%	0,568
GO:0048728	proboscis development	0.000%	0,565
GO:0008407	chaeta morphogenesis	0.002%	0,565
GO:0045468	regulation of R8 cell spacing in compound eye	0.000%	0,578
GO:0060729	intestinal epithelial structure maintenance	0.002%	0,604
GO:0016183	synaptic vesicle coating	0.000%	0,838
GO:0007295	growth of a germarium-derived egg chamber	0.001%	0,562
GO:0035263	genital disc sexually dimorphic development	0.000%	0,529
GO:0045201	maintenance of neuroblast polarity	0.000%	0,569
GO:0007406	negative regulation of neuroblast proliferation	0.003%	0,508
GO:0042078	germ-line stem cell division	0.003%	0,533
GO:0048133	male germ-line stem cell asymmetric division	0.001%	0,55

Continued on next page

Table S6 – continued from previous page

GO term ID	Description	Freq.	Uniq.
GO:0051572	negative regulation of histone H3-K4 methylation	0.001%	0,807
GO:0031936	negative regulation of chromatin silencing	0.010%	0,728
GO:0051573	regulation of histone H3-K9 methylation	0.003%	0,801
GO:0071441	negative regulation of histone H3-K14 acetylation	0.003%	0,802
GO:0048821	erythrocyte development	0.007%	0,5
GO:0007561	imaginal disc eversion	0.000%	0,543
GO:0008354	germ cell migration	0.004%	0,555
GO:0007413	axonal fasciculation	0.006%	0,506
GO:0016199	axon midline choice point recognition	0.002%	0,495
GO:0016203	muscle attachment	0.003%	0,576
GO:0048636	positive regulation of muscle organ development	0.012%	0,512
GO:0008258	head involution	0.002%	0,586
GO:0007391	dorsal closure	0.005%	0,538
GO:0046928	regulation of neurotransmitter secretion	0.012%	0,783
GO:0007500	mesodermal cell fate determination	0.000%	0,538
GO:0007501	mesodermal cell fate specification	0.004%	0,497
GO:0007509	mesoderm migration involved in gastrulation	0.001%	0,527
GO:0035309	wing and notum subfield formation	0.001%	0,53
GO:0007475	apposition of dorsal and ventral imaginal disc-derived wing surfaces	0.001%	0,525
GO:0002121	inter-male aggressive behavior	0.004%	0,899
GO:0042683	negative regulation of compound eye cone cell fate specification	0.000%	0,535
GO:0009997	negative regulation of cardioblast cell fate specification	0.001%	0,525
GO:0035073	pupariation	0.001%	0,594
GO:0016348	imaginal disc-derived leg joint morphogenesis	0.001%	0,501
GO:0035317	imaginal disc-derived wing hair organization	0.002%	0,451
GO:0008587	imaginal disc-derived wing margin morphogenesis	0.002%	0,507
GO:0008586	imaginal disc-derived wing vein morphogenesis	0.002%	0,508
GO:0002168	instar larval development	0.003%	0,573
GO:0048800	antennal morphogenesis	0.001%	0,527

Continued on next page

Table S6 – continued from previous page

GO term ID	Description	Freq.	Uniq.
GO:0007480	imaginal disc-derived leg morphogenesis	0.004%	0,496
GO:0001737	establishment of imaginal disc-derived wing hair orientation	0.001%	0,461
GO:0007474	imaginal disc-derived wing vein specification	0.003%	0,485
GO:0007455	eye-antennal disc morphogenesis	0.002%	0,512
GO:0007447	imaginal disc pattern formation	0.005%	0,498
GO:0051384	response to glucocorticoid	0.012%	0,921
GO:0051315	attachment of mitotic spindle microtubules to kinetochore	0.011%	0,751
GO:0007638	mechanosensory behavior	0.004%	0,846
GO:0050974	detection of mechanical stimulus involved in sensory perception	0.007%	0,617
GO:0051591	response to cAMP	0.010%	0,919
GO:0007291	sperm individualization	0.003%	0,518
GO:0031987	locomotion involved in locomotory behavior	0.005%	0,922
GO:0046845	branched duct epithelial cell fate determination, open tracheal system	0.000%	0,587
GO:0048854	brain morphogenesis	0.010%	0,536
GO:0008355	olfactory learning	0.005%	0,569
GO:0007632	visual behavior	0.012%	0,559
GO:0016476	regulation of embryonic cell shape	0.002%	0,516
GO:0042381	hemolymph coagulation	0.001%	0,599
GO:0007141	male meiosis I	0.004%	0,732
GO:0000712	resolution of meiotic recombination intermediates	0.008%	0,709
GO:0007111	meiosis II cytokinesis	0.001%	0,742
GO:0007110	meiosis I cytokinesis	0.001%	0,74
GO:0007112	male meiosis cytokinesis	0.002%	0,729
GO:0071312	cellular response to alkaloid	0.005%	0,908
GO:0031290	retinal ganglion cell axon guidance	0.006%	0,469
GO:0008045	motor neuron axon guidance	0.011%	0,458
GO:0048841	regulation of axon extension involved in axon guidance	0.004%	0,435
GO:0045463	R8 cell development	0.000%	0,555
GO:0045467	R7 cell development	0.001%	0,534
GO:0045470	R8 cell-mediated photoreceptor organization	0.000%	0,548
GO:0007460	R8 cell fate commitment	0.000%	0,536
GO:0007425	epithelial cell fate determination, open tracheal system	0.000%	0,578

Continued on next page

Table S6 – continued from previous page

GO term ID	Description	Freq.	Uniq.
GO:0009950	dorsal/ventral axis specification	0.008%	0,545
GO:0021513	spinal cord dorsal/ventral patterning	0.005%	0,529
GO:0034059	response to anoxia	0.002%	0,936
GO:0051533	positive regulation of NFAT protein import into nucleus	0.002%	0,837

Table S7: Differential expression data of opsin and cryptochrome genes in the transcriptomic experiment. LD vs SD_{14} photoperiods. q-value is an adjusted p-value considering false discovery rate (FDR).

Gene	ACYPI	FPKM LD	FPKM SD_{14}	$\log_2$(fold change)	test stat	q-value
Ap-LWO	ACYPI009332	99,41	117,47	0,240	0,337	0,999
Ap-SWO1	ACYPI001006	5,06	4,29	-0,239	-0,372	0,999
Ap-SWO2	ACYPI004442	11,51	13,37	0,216	0,364	0,999
Ap-SWO3	ACYPI002544	17,61	17,43	-0,015	-0,027	0,999
Ap-SWO4	ACYPI005074	0,53	0,80	0,586	0,692	0,999
Ap-arthropsin	ACYPI32968	16,69	20,74	0,313	0,557	0,999
Ap-C-Ops	ACYPI009397	2,45	3,03	0,303	0,092	0,999
Cry 1	ACYPI005768	2,69	2,57	-0,066	-0,101	0,999
Cry 2-1	ACYPI006584	0,012	0,01	0,117	0	1

Table S8: Differential expression data of opsin and cryptochrome genes found in transcriptomic experiment. LD vs SD_{10}. Significant values in bold. q-value is an adjusted p-value considering false discovery rate (FDR).

Gene	ACYPI	FPKM LD	FPKM SD_{10}	$\log_2$(fold change)	test stat	q-value
Ap-LWO	ACYPI009332	90,16	127,86	0,504	0,687	0,567
Ap-SWO1	ACYPI001006	4,60	4,94	0,102	0,153	0,955
Ap-SWO2	ACYPI004442	10,44	13,09	0,326	0,549	0,701
Ap-SWO3	ACYPI002544	15,98	23,58	0,560	0,984	0,338
Ap-SWO4	ACYPI005074	0,48	0,93	0,948	1,074	0,286
Ap-Arthropsin	ACYPI32968	15,14	31,26	1,046	1,840	**0,026**
Ap-C-Ops	ACYPI009397	2,23	3,18	0,512	0,124	0,960
Cry 1	ACYPI005768	2,44	3,06	0,329	0,494	0,748
Cry 2-1	ACYPI006584	0,01	0	-inf	0	1

Table S9: List of primers used for PCR amplification, sequencing, RT-qPCR and riboprobe synthesis of the genes characterised in the present report.

Gene Name*	Primer Name	Sequence (5' → 3')
Ap-LWO	ApLwoF1	GTACTACTTCAATATCTGACAGTC
	ApLwoF2†	TGGTCACATTCTACTGGAGCATC
	ApLwoR1	TCATATCAAAGATACGCTAGAGGC
	OpsCiF§	GATGGAGCCGATATGTACCAGAA
	OpsCiR§†	AATCAGTGCCACATGCTGTCAT
Ap-SWO1	Uv1op-FQ§†	CGCTAGGAAAATTGGGATGCCAG
	Uv1op-RQ§	GCTTTTAGAGGGCACACTATAAC
	Uv1op-F1	GCCTCTTATCGTTGCTGAGGTTC
	Uv1op-R1†	TTTAGTTCTTCAATGATTTATTC
	Uv1op-F2	TTATCGTTGCTGAGGTTCAATTC
	Uv1op-F3†	GCTGTGGTTTGTGTCGTGGACGC
	Uv1op-R2†	CGAAGACACCCAGCAGTGCCACC
Ap-SWO2	Uv2op-FQ§	GCACTTCGTGAACAAGCTAAAAAG
	Uv2op-RQ§†	GCAGCTACAAACAGGAAACAAATC
	Uv2op-F1†	CGTTTGTTTCGGTCGTAGTAGTG
	Uv2op-R1	TATATTCGGTATATAACTTTCTG
	Uv2op-F2	TTCGGTCGTAGTAGTGACTAGTG
Ap-SWO3	Uv3op-FQ§	TCTATAAATCGTGGTTATCAGGG
	Uv3op-RQ§†	ATGGTTTTGCTATTGTACTAAATC
	Uv3op-F1	ATATACCTCCCCACAGCCCTTAC
	Uv3op-R1	TGTATATTCTTTGGAGTAGACAG
	Uv3op-R2	AGACACACGTTCAGTTATTTATC
	Uv3op-R3	ATTGTACTAAATCTGTCGTATGC
	Uv3op-F2	CTCCCCACAGCCCTTACGTACATC
	Uv3op-F3†	CGCGTACACAGCCAGTCCAC
Ap-C-Ops	ApPops-RQ§	CGTTTTGAGGAGTCCATAGCCTTG
	ApPops-FQ§	CACATTTCAGATGCCATCTACC
	ApPops-F2†	TTGATGTCAAACATTCAATGTGG
	ApPops-R1†	TGTTTTGTTTCGTTTCAACGGTTG
	ApPops-F1	AAATACGTTGATGTCAAACATTC
	ApPops-F3†	CGTCGTTACAGTTAGAAAGAGAG
	ApPops-R2†	ATAATGGCTACCCTGCATTCTGC
Ap-SWO4	ApMops-FQ§	GTACTCGGTTGTTTCGGGAACTC
	ApMops-RQ§†	CCATAGATCTGGCAGCCCAATTTTC
	ApMops-F2†	CAACTATGAACAGTCCATCAGCC
	ApMops-R1	AAATCTTCAAACTCATCGTGGGC
	ApMops-F1	ACCGTTGTTTTTGCGACTGTCTTG
	ApMops-F3†	ACCTGTCGGATGACAATCAAGAG
	ApMops-R2†	TCTTCATACCATAATCCCGCTCC
	ApMops-R3	CAAACTCATCGTGGGCTCGTTGGC
Ap-Arth.	Artrops-F1	GGGTAAAAGACCAGAATCATCG

Artrops-F2	GGAGTACCTATACCTACTGATG
Artrops-F3†	AATGGAAAACTGGGCGAATGAG
Artrops-F4Q§	TTGGGCACTGAGAGTTGGGA
Artrops-R1	TTACATACTGACCCTGGTCAGTC
Artrops-R2	CATTGTAGGTACCTACCTACAC
Artrops-R4Q§†	ACCACAGACGCCTTAGCGCA
ArtrM-R6†	TATAAATCCGAGCACCAACATG
Artrhops-F5†	ACTCTTTCCACCATAGCGATTG

*Gene name: See text for details on gene names (section 3.3.3)
§Primers used for RT-qPCR
†Primers used for riboprobe synthesis

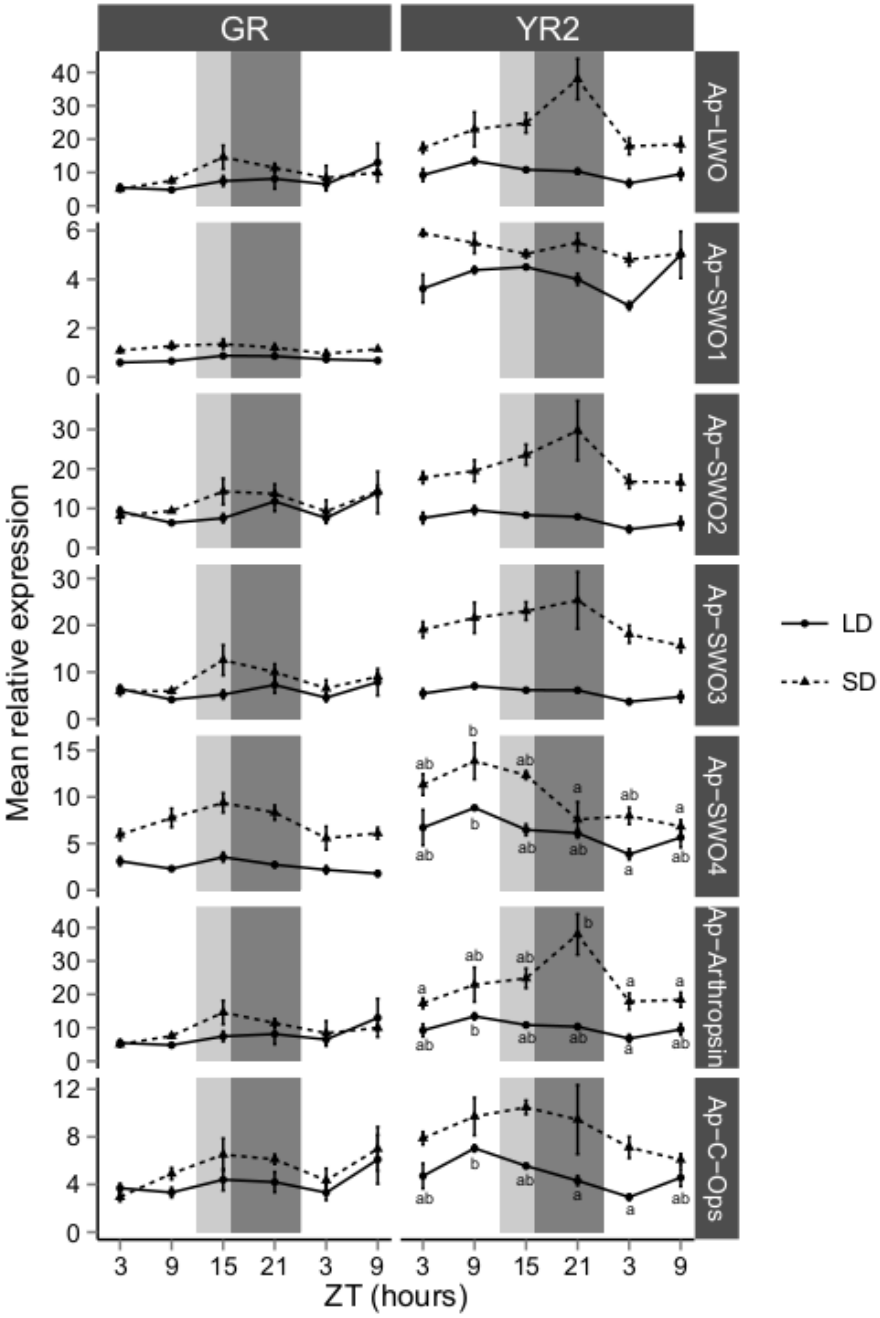

GR
YR2
Ap−LWO
Ap−SWO1
Ap−SWO2
Ap−SWO3
Ap−SWO4
Ap−Arthropsin
Ap−C−Ops
Mean relative expression
ZT (hours)
LD
SD

Figure S2: Previous page: Relative expression of opsin genes of *A. pisum* along 1,5 day/night cycles. Dark phases are shaded. Error bars represent the standard error of the mean (SEM). Homogeneous subgroups are labelled a; b; ab when significant differences were found (Tukey HSD, $P<0.05$). Ap-LWO, *A. pisum* Long Wavelength Opsin; Ap-SWO1 to Ap-SWO4, *A. pisum* Short Wavelength Opsins 1 to 4; Ap-Arthropsin, *A. pisum* Arthropsin; Ap-C-Ops, *A. pisum* C-opsin; GR, Gallur Red; YR2, York Red; ZT, zeitgeber; SD, Short day; LD, Long day.

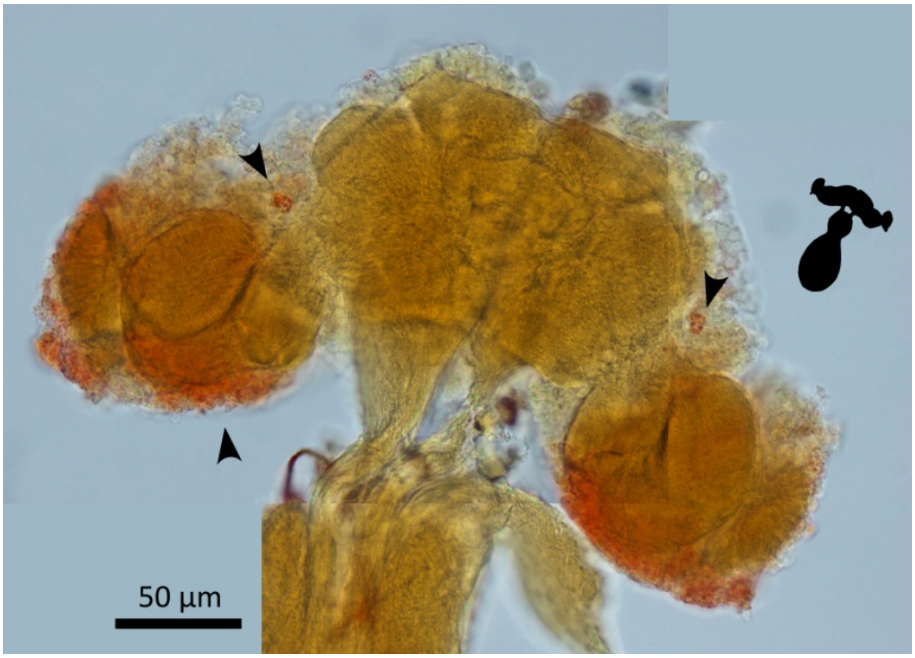


Figure S3: *In situ* localisation of Ap-Arthropsin gene transcripts in adult *A. pisum* central nervous system. Arrowheads point to specifically stained regions. A schematic representation of the aphid central nervous system is included beside each image indicating the orientation of the preparation. Riboprobes specifically bound are detected with Fast Red/HNPP (see Materials and Methods)

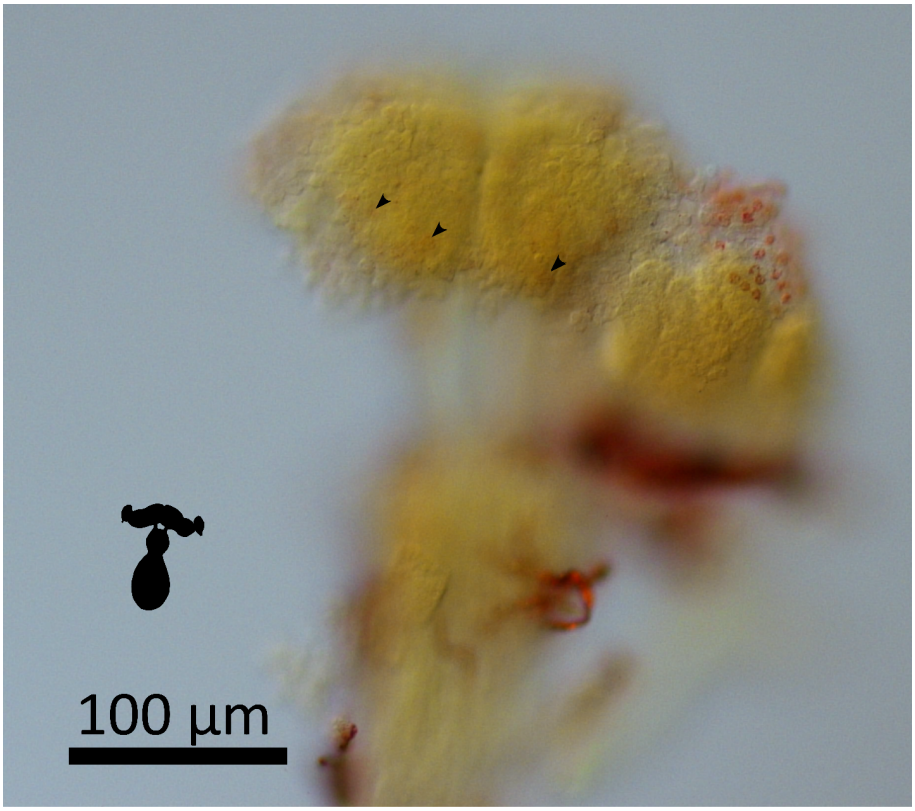


Figure S4: *In situ* localisation of Ap-C-ops gene transcripts in adult *A. pisum* central nervous system. Arrowheads point to specifically stained regions. A schematic representation of the aphid central nervous system is included beside each image indicating the orientation of the preparation. Riboprobes specifically bound are detected with Fast Red/HNPP (see Materials and Methods)

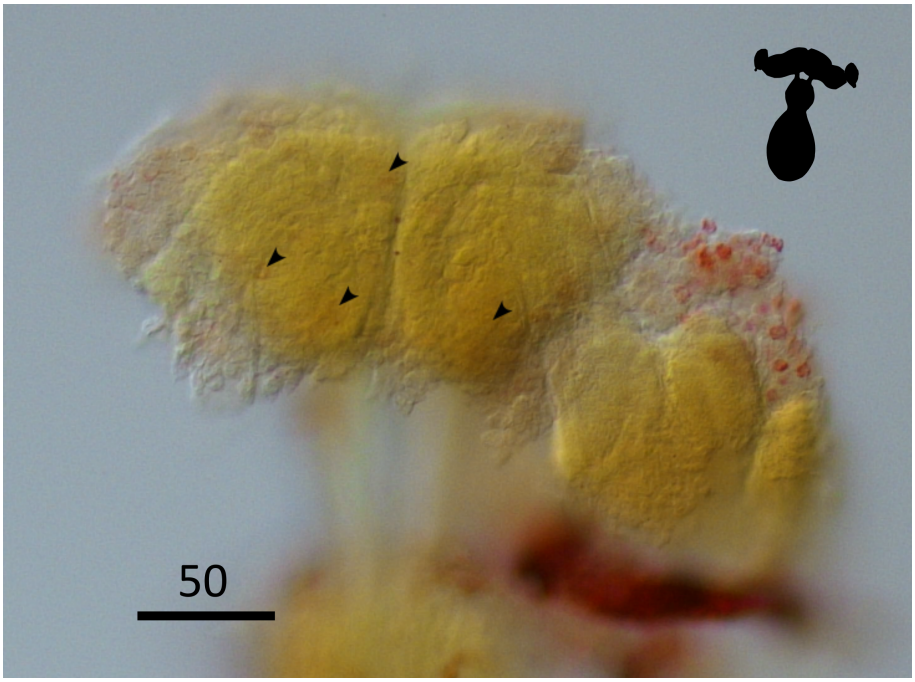


Figure S5: *In situ* localisation of Ap-C-ops gene transcripts in adult *A. pisum* central nervous system. This is an amplification of a differ-ent focus distance of Supplementary Figure S3. Arrowheads point to specifically stained regions. A schematic representation of the aphid central nervous system is included beside each image indicating the orientation of the preparation. Riboprobes specifically bound are detected with Fast Red/HNPP (see Materials and Methods)