Insect Molecular Biology

Insect Molecular Biology (2010), **19** (Suppl. 2), 123–139 doi: 10.1111/j.1365-2583.2009.00931.x

Identification and characterization of circadian clock genes in the pea aphid *Acyrthosiphon pisum*

T. Cortés, B. Ortiz-Rivas and D. Martínez-Torres

Institut Cavanilles de Biodiversitat i Biologia Evolutiva, Universitat de València, Valencia, Spain

Abstract

The molecular basis of circadian clocks is highly evolutionarily conserved and has been best characterized in *Drosophila* and mouse. Analysis of the *Acyrthosiphon pisum* genome revealed the presence of orthologs of the following genes constituting the core of the circadian clock in *Drosophila*: *period* (*per*), *timeless* (*tim*), *Clock*, *cycle*, *vrille*, and *Pdp1*. However, the presence in *A. pisum* of orthologs of a mammal-type in addition to a *Drosophila*-type *cryptochrome* places the putative aphid clockwork closer to the ancestral insect system than to the *Drosophila* one. Most notably, five of these putative aphid core clock genes are highly divergent and exhibit accelerated rates of change (especially *per* and *tim* orthologs) suggesting that the aphid circadian clock has evolved to adapt to (unknown) aphid-specific needs. Additionally, with the exception of *jetlag* (absent in the aphid) other genes included in the *Drosophila* circadian clock repertoire were found to be conserved in *A. pisum*. Expression analysis revealed circadian rhythmicity for some core genes as well as a significant effect of photoperiod in the amplitude of oscillations.

Keywords: aphid, photoperiod, clock genes, circadian expression.

Introduction

Circadian clocks are internal endogenous oscillators governing daily cycles of activity in most organisms including their physiology and behaviour (Bell-Pedersen *et al.*, 2005). Photoperiod is the main stimulus entraining or synchronizing the clock to match the day–night cycles and, since it is also the main cue for season change, the participation of the circadian clock to keep seasonal rhythms has also been suggested (Lincoln *et al.*, 2003; Stoleru *et al.*, 2007).

Basic molecular constituents of animal circadian clocks are highly evolutionarily conserved and have been best characterized in *Drosophila* and mouse. In *Drosophila*, six transcription factors, organized into two feedback loops (Fig. 1), are central to the circadian clock (Cyran *et al.*, 2003; Hardin, 2005). The *per/tim* feedback loop is based on the rhythmic expression of the *period* (*per*) and *timeless* (*tim*) genes, which encode for the transcription factors PER and TIM which repress their own transcription. Expression of *per* and *tim* is promoted by CLOCK-CYCLE (CLK-CYC) heterodimers binding E-box promoter regulatory elements. Levels of *per* and *tim* transcripts peak at dusk but PER and TIM accumulate in a delayed fashion during the night forming PER-TIM complexes that can enter into the nucleus. Once in the nucleus, PER interacts with CLK-CYC heterodimers bound to E-boxes, which inhibits CLK-CYC-mediated transcriptional activation, leading to minimum levels of *per* and *tim* mRNAs at dawn (Hardin, 2005). The kinases *double-time* (DBT), *Casein kinase 2* (CK2) and *shaggy* (SGG), the *protein phosphatase 2a* (PP2A) and the F-box degradation signalling protein *supernumerary limbs* (SLMB) are involved in regulating the stability, subcellular localization and eventual degradation of PER and TIM proteins. In a second feedback loop (the *Clk* loop), CLK-CYC heterodimers directly activate at dusk the transcription of genes v*rille* and *Pdp1* encoding the transcription factors VRI and PDP1 (specifically, the PDP1epsilon isoform) which inhibit and activate, respectively, the expression of *Clock* so that peak levels occur approximately in anti-phase to that of *per* and the other CLK-CYC-activated genes (Cyran *et al.*, 2003). Photoperiodic entrainment is accomplished by the light-induced degradation of TIM through the participation of the photoreceptor CRYPTOCHROME (CRY1 in Fig. 1) and the F-box E3 ubiquitin ligase JETLAG (Koh *et al.*, 2006; Peschel *et al.*, 2006).

Correspondence: David Martínez-Torres, Institut Cavanilles de Biodiversitat i Biologia Evolutiva, Universitat de València, Apartado de Correos 22085, 46071 Valencia, Spain. Tel.: +34 96354 3644; fax: +34 96354 3670; e-mail: david.martinez@uv.es

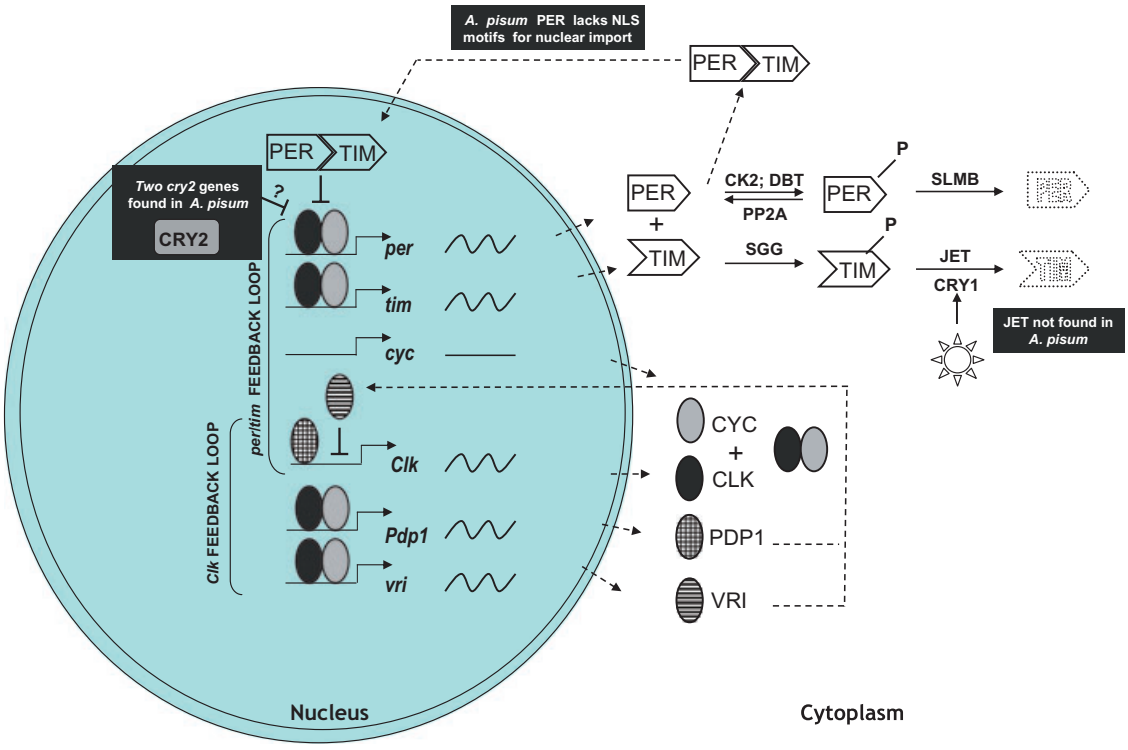

Figure 1. Model of the circadian clock in *Drosophila*. Genes (represented by crooked arrows) are organized into two interconnected feedback loops (Cyran *et al*., 2003) including the orthologs present in *Acyrthosiphon pisum* and major differences found in the aphid are highlighted in black boxes. Proteins constituting the core of the two feedback loops in *Drosophila* are indicated by different shapes. See text for details. Proteins committed to degradation are indicated by dotted shapes. The pea aphid genome contains two copies of a mammalian-type cryptochrome, CRY2, which is absent in *Drosophila*. Their 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 *Drosophila*.

Two transcriptional feedback loops similar to those just described for *Drosophila* are also at the core of the circadian clockwork in the mouse, with conserved orthologs of most of the genes involved but having several differences. First, multiple copies of several clock genes occur in mammals leading to increased complexity. Second, orthologs of both *timeless* and the *Drosophila*-type *cryptochrome* are absent and there is an inclusion of two paralogs of an evolutionarily distinct *cryptochrome* (a so-called mammalian-type *cryptochrome* or *mCry1* and *mCry2*, reviewed in Looby & Loudon, 2005). In the *Drosophila* central pacemaker, *cryptochrome* synchronizes the *per/tim* feedback loop with the light–dark cycle, whereas mCRYs are an integral part of the clockwork, substituting for TIM as the partner/s of PER to repress CLK-BMAL heterodimer activity (*Bmal* is the mammalian ortholog of *cycle*; Ivanchenko *et al*., 2001; Lin & Todo, 2005; Looby & Loudon, 2005; see also Collins *et al*., 2006, for a repressor role of *Drosophila* CRY in peripheral clocks). In mammals, the second loop works similarly to the *Clk* loop described for *Drosophila*, but cyclic expression of the activator *Bmal* instead of *Clk* is controlled by the products of two genes, *Rev-Erbα* and *Rora* not related with the *Drosophila vrille* and *Pdp1* genes (reviewed in Bell-Pedersen *et al*., 2005; Looby & Loudon, 2005). As a net result of cyclic activity of clock genes, oscillations in the level of expression or in the activity of clock-controlled genes finally lead to the overt biological rhythms.

Given this high degree of conservation between *Drosophila* and mammals, it was expected that the *Drosophila* clock model would be highly conserved among other insects (Rubin *et al*., 2006). However, current data derived from circadian specific studies and from the analysis of recently available insect genomes suggest that, contrary to expectations, the *Drosophila* model cannot be generalized to the rest of the insect species and that diverse clock mechanisms may exist among them (Yuan *et al*., 2007). Differences in localization, pattern of expression, gene structure and in the gene repertoire involved in the clockwork have been reported (Sauman & Reppert, 1996; Chang *et al*., 2003; Zhu *et al*., 2005, 2008; Rubin *et al*., 2006; Yuan *et al*., 2007; Ikeno *et al*., 2008). More importantly, a mammalian-type *cryptochrome*, designated insect CRY2, that is absent in *Drosophila* is present in all non-drosophilid insects studied to date. Moreover, all insect CRY2 proteins potently repress CLK-CYC

transcriptional activation *in vitro*, a property that is shared with mCRYs (Yuan *et al*., 2007). The *Drosophila*-type *cryptochrome* (here designated insect CRY1 as proposed by Yuan *et al*., 2007) is absent from the genomes of *Tribolium castaneum* and *Apis mellifera* and the latter species also lacks a *timeless* ortholog in its genome. Considering all these elements together, three major types of clockwork model have been proposed in insects (Yuan *et al*., 2007). First, the previously described *Drosophila* type in which PER would be the main repressor of CLK-CYC and CRY1 would merely transmit photic information to the clockwork through TIM. Second, the butterfly model having both CRY1 and CRY2, but the latter instead of PER would act as the main repressor of CLK-CYC transcriptional activation. Third, the absence of CRY1 in *Tribolium* and *Apis* depicts a clockwork type where, as in the butterfly, the main repressing function would reside in CRY2 which could also have a role in light entrainment.

Aphids, like most organisms, exhibit diurnal changes in their physiology and behaviour which are likely to be controlled by an endogenous circadian clock. Whilst limited, there are examples reporting the daily rhythm of release of sex pheromones (Eisenbach & Mittler, 1980; Thieme & Dixon, 1996), the rhythm in fresh weight-gain and larviposition time (Hodgson & Lane, 1981) or the rhythm in host-finding behaviour (Narayandas & Alyokhin, 2006). However, there is one aspect of aphid biology, related to the circadian clock, that has received much attention and that is their mode of reproduction by cyclical parthenogenesis. The switch from a parthenogenetic to a sexual mode of reproduction (or its inhibition by manipulations of the light–dark cycles) constitutes the first case of photoperiodism studied in animals (Markovitch, 1924) and has since been one of the most thoroughly studied examples of photoperiodism in insects. Briefly, shortening of day-length in autumn is the main cause of the switch from viviparous parthenogenesis to oviparous sexual reproduction in aphids living in nature. Moreover, this process can be mimicked or prevented in the laboratory by convenient manipulations in the photoperiod experienced by the reared aphids (Hardie & Vaz Nunes, 2001). Aphids had long been considered to measure photoperiodic time by a non-oscillatory (non-circadian) or hourglass mechanism (Hillman, 1973; Lees, 1973). Current evidence, however, suggests that photoperiodic responses in insects, including aphids, as in other major taxa are a function of the circadian system (Vaz Nunes & Hardie, 1993; Hardie & Vaz Nunes, 2001; Saunders *et al*., 2004; but see Veerman, 2001; Bradshaw *et al*., 2003; Emerson *et al*., 2009). Determining the existence and analysing the expression of known clock genes in species with robust photoperiodic responses is an essential step towards elucidation of this important issue. The recent availability of the genome of the aphid *Acyrthosiphon pisum* by the International Aphid Genomics Consortium offers the opportunity not only to characterize the first clockwork system in a hemimetabolous insect but also in a robustly photoperiodic system. This will allow a better understanding of the evolution of insect circadian clocks (current models are built based only on holometabolous species) and will elucidate whether the circadian clockwork (or some of its elements) also participates in the photoperiodic response.

In this report we present results of a search through the *A. pisum* genome for homologs of genes known to participate in the circadian clocks of *Drosophila* and other insects in order to establish the gene repertoire constituting the aphid clockwork. Using genome information, we cloned and sequenced aphid cDNAs from the genes constituting the core of the two feedback loops described above including *per* and *tim* homologs, for which predictions were rather poor, but also *Clk*, c*yc*, *Cry1*, *Cry2*, v*ri* and *Pdp1*. For genes involved in stability/degradation of clock proteins present in the aphid genome, models available from the gene prediction programs, were either validated or modified by comparison with sequences from other insect species. Phylogenetic analysis was also carried out both to ensure orthology of the sequences and to better understand the evolution of clock genes. Finally, we investigated the expression of the eight core genes along the day–night cycle under two different photoperiod conditions.

Results

Identification of clock gene homologs in Acyrthosiphon pisum

Homologs for most of the *Drosophila* clock genes (Fig. 1) were identified in the *A. pisum* genome (Table 1). Most of these genes were found among the set of NCBI's RefSeq predictions. However, two key genes in the *Drosophila* clock (*period* and *timeless*) were absent from this set. For these two genes no expressed sequence tags (ESTs) were present in the databases (Table 1) and only partial predictions were available among non-RefSeq gene models included in the GLEAN set (Elsik *et al*., 2007) (Table 1). For *AcpPer* two poorly predicted non-overlapping models were available on two different scaffolds, adding up to 1069 translated amino acids. In the case of *AcpTim* a single model that could be translated into 477 amino acids was available (Table 1). A *timeless* paralog (a true ortholog of the *Drosophila Timeout* gene, whose involvement in the clockwork system is unknown), was, however, found among RefSeq predictions (see below). Both *Drosophila*-type (*AcpCry1*) and mammalian-type (*AcpCry2*) *cryptochrome* genes were identified in the aphid genome. Interestingly, the *AcpCry2* gene was also

Table 1. Summary of data obtained for clock gene sequences identified in the *Acyrthosiphon pisum* genome

Genes[1]	Symbol[2]	Prediction[3]	Aphidbase[4]	ESTs[5]	NCBI[6]	Length[7]	Comp.[8]	Rate[9]
Feedback loop core genes								
period	*AcpPer*	GLEAN_27669	ACYPI47669	0	FM998646	1018 / 462	0,009**	++
		GLEAN_36135	ACYPI48687					
timeless	*AcpTim*	GLEAN_16439	ACYPI36439	0	FM998651$	606 / 382	0,006**	++
cycle	*AcpCyc*	XP_001943647	ACYPI004686	2	FM998648	648 / 273	0,519	+ (4)
clock	*AcpClk*	XP_001944549	ACYPI004812	3	FM998647	613 / 323	0,983	+ (2)
vrille	*AcpVri*	XP_001949829	ACYPI008851	0	FM998650$	485 / 148	1,000	=
PAR-domain protein 1	*AcpPdp1*	XP_001943257	ACYPI000702	4	FM998649	293 / 147	1,000	=
cryptochrome	*AcpCry1*	XP_001944402	ACYPI005757	2	FN377569	557 / 466	0,447	=
	AcpCry2-1	XP_001950693	ACYPI006584	0	FN377570	501 / 491	0,830	+ (4)
	AcpCry2-2	XP_001950192	ACYPI004197	0	FN377571	491 / 491	0,664	+ (4)
Stability/degradation								
double-time	*AcpDbt*	XP_001951697	ACYPI008162	10	–	417 / 261	1,000	=
supernumerary limbs	*AcpSlmb*	XP_001949585	ACYPI006874	35	–	525 / 418	1,000	=
Casein kinase II α subunit	*AcpCk2α*	XP_001942962	ACYPI002006	33	–	361 / 322	1,000	=
Casein kinase II β subunit	*AcpCk2β*	NP_001119697	ACYPI000089	26	–	220 / 219	1,000	=
protein phosphatase 2a (twins)	*AcpTws*	XP_001949195	ACYPI009741	14	–	448 / 440	1,000	=
protein phosphatase 2a (widerborst)	*AcpWdb*	XP_001944239	ACYPI000666	20	–	579 / 428	1,000	=
shaggy	*AcpSgg*	XP_001951519	ACYPI009257	11	–	425 / 349	1,000	=
jetlag	Not found							

[1]*Drosophila* genes searched for in the *A. pisum* genome.
[2]Symbols given to the identified genes.
[3]If available, NCBI RefSeq accession number of the predicted orthologous protein is given. Otherwise available GLEAN models are given.
[4]Aphidbase Gbrowse accession numbers (http://www.aphidbase.com/aphidbase/).
[5]Number of ESTs found among the approximately 170.000 *A. pisum* ESTs available at Aphidbase.
[6]NCBI accession numbers of cDNAs sequenced in the present study ($, partial sequences).
[7]Length of the *A. pisum* gene sequences (amino acids)/Length of the alignments of conserved regions used for phylogenetic reconstructions. For the core genes (except *AcpVri*), the length derived from our sequenced cDNAs is given.
[8]p-values obtained for *A. pisum* sequences in the chi-square tests performed using TREE-PUZZLE (Schmidt *et al.*, 2002) to test for homogeneity of amino acid composition in insect sequences. **, highly significant ($P < 0.01$).
[9]Summary of the results obtained using RRTree (Robinson-Rechavi & Huchon, 2000) to test for homogeneity in rates of amino acid sequence evolution among insect sequences. ++, *A. pisum* sequences showing highly accelerated rates in all the comparisons; +, *A. pisum* sequences showing accelerated rates in some of the comparisons (the number of significant comparisons over a total of 5 is indicated in parenthesis); =, *A. pisum* sequences not showing accelerated rates in any comparison. Insect species included in the composition and rate analyses are as in Fig. 2.

duplicated. Seven of the eight genes putatively involved in the stability/degradation of core clock proteins were also searched for, and found to be present in the aphid genome (Table 1). However, our searches failed to identify a true ortholog of *Drosophila* E3 ubiquitin ligase *jetlag*. Sequence analysis, including alignments of the predicted genes with available sequences from different species and phylogenetic analysis, allowed us to confirm the true orthology of the sequences indicated in Table 1 and to validate the RefSeq models for the most part. All the identified genes were manually annotated in the Apollo GMOD-Chado database accessible at the AphidBase web portal (www.aphidbase.com). Except for *period* and *timeless*, only a few gene models needed minor modifications. Additionally, we designed primers based on the genome sequence (Table S1), and cDNAs from the core clock gene set were amplified by PCR and sequenced. This allowed experimental validation of the corresponding gene models and was essential for obtaining the true coding sequence of *A. pisum* genes *period* and *timeless* (Table 1 and below). Finally, the numbers of ESTs available for putative core clock genes were very low, ranging from none to four, while those available for genes involved in the stability/degradation of core clock genes were more abundant (Table 1).

Evolutionary distinctiveness of clock genes in Acyrthosiphon pisum

During the course of sequence analysis it soon became evident that, although some aphid clock gene sequences were readily aligned with orthologous sequences from other insect species, the alignments contained a high number of mismatches. Sequencing errors were excluded after examining traces available at NCBI for some of these genes and after cDNA sequencing of others (see Experimental procedures). Homogeneity tests performed with TREEPUZZLE (Schmidt *et al.*, 2002, see Experimental procedures) on available insect sequences, representative of different insect orders, including the hemipteroid *Pediculus humanus*, revealed significant differences in amino acid composition in *A. pisum* AcpPER and AcpTIM (Table 1). All other insect species showed homogeneous composition for all sequences analysed. Moreover,

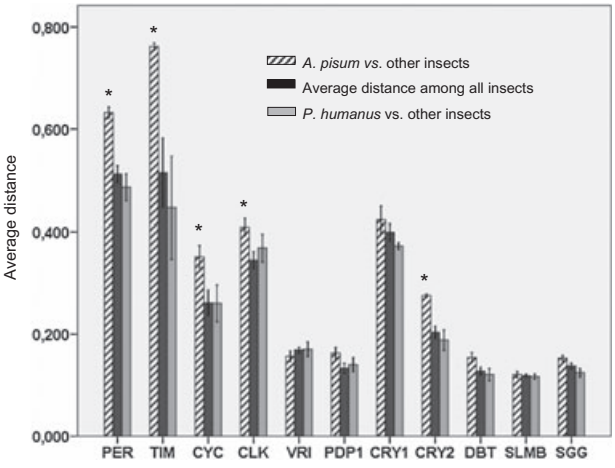

Figure 2. Average amino acid distances ± SE for several clock genes. Striped bars represent average distances between *Acyrthosiphon pisum* and other insects belonging to different orders including a second hemipteroid, *Pediculus humanus* (Anoplura), from the Hemimetabola and *Drosophila melanogaster* (Diptera), *Bombyx mori* (Lepidoptera), *Apis mellifera* (Hymenoptera) and *Tribolium castaneum* (Coleoptera) from the Holometabola. Black bars show the average distances for each gene among all the above-mentioned species (*A. pisum* included). Grey bars represent average distances between the hemipteroid *P. humanus* and the rest of insect species, including *A. pisum*. The orthologous sequences of PER and CRY2 from *Riptortus pedestris* (Hemiptera) and of PER from *Blattela germanica* (Orthoptera) were also included in the corresponding analysis. Due to problems with some sequences retrieved from databases, or to the lack of genes in some species, *Anopheles gambiae* (Diptera) replaced *D. melanogaster* in CRY2 and CYC analyses, *Dianemobius nigrofasciatus* (Orthoptera) replaced *P. humanus* in CRY1 analysis and *Antheraea pernyi* and *Danaus plexippus* (Lepidoptera) replaced *B. mori* in CLK and SGG analyses, respectively. Significant difference ($P < 0.01$), as revealed by one-way ANOVA comparing bar heights, were due to extreme aphid divergence (indicated by asterisks).

relative rate tests performed on amino acid sequences using RRTREE (Robinson-Rechavi & Huchon, 2000 see Experimental procedures), showed that AcpPER, AcpTIM, and other genes orthologous to those participating in the *Drosophila per*/*tim* feedback loop (Fig. 1) were evolving at accelerated rates with respect to their respective insect orthologs (Table 1). Additionally, by computing the percentage of amino acid differences among available clock gene sequences representative of different insect orders and comparing with the average distance between *A. pisum* and the rest of insect sequences, it was evident that some aphid clock genes were particularly divergent (Fig. 2). This unusual divergence was not observed, however, between equivalent sequences from a second hemipteroid (*P. humanus*) or in any other insect sequences. As a result, some aphid sequences, although clearly orthologous of corresponding insect genes, occupied aberrant positions in the phylogenetic trees (some of them are shown in the next section). These results support the hypothesis that genes involved in the putative *per*/*tim* feedback loop (Fig. 1), especially *period* and *timeless*, are evolving faster in *A. pisum* than their orthologs in other insects. However, putative genes from the *Clk* feedback loop and genes involved in the regulation of the stability and degradation of core clock proteins (Fig. 1) appear to have evolved at rates comparable to other insect species (Table 1; Fig. 2).

Characterization of Acyrthosiphon pisum *core-clock gene sequences period and timeless*

Cloning and sequence analysis of *AcpPer* cDNA yielded a 3275 bp sequence that contained both an initial AUG and a stop codon and a predicted protein of 1018 amino acids. The predicted protein clearly grouped with other insect PER sequences after a phylogenetic analysis that revealed, however, a high divergence of the *A. pisum* sequence (Figs. 2, 3A; Table 1). As a PAS protein (PER-ARNT-SIM; Hirayama & Sassone-Corsi, 2005), AcpPER was shown, after a SMART search, to contain two tandemly organized PAS (PAS-A and PAS-B) domains and a PAC (PAS associated C-terminal) domain highly conserved among all animal PER proteins (Fig. 3B). While these three domains are recognized in *A. pisum*, they are rather divergent. For instance, two regions corresponding to proposed TIM binding sites in *Drosophila* were found to be highly divergent, as well as a region orthologous to the CLK-CYC interaction domain or CCID (Fig. 3B; Saez & Young, 1996; Chang & Reppert, 2003). Similarly, two Nuclear Localization Signal (NLS) motifs necessary for nuclear import of PER that are well conserved across insect species are either missing or highly modified in AcpPER (Fig. 3B).

PCR amplification of *AcpTim* cDNA usually needed two rounds of nested PCR reactions and, in many instances, fragments apparently containing unspliced introns were obtained. We PCR amplified, cloned and sequenced 1818 bp of cDNA corresponding to a partial sequence of the *A. pisum timeless* homolog encoding 606 amino acids (about half the size of the protein in other insects) with the 5′- and 3′- ends missing (about 70 and 350–400 amino acids, respectively, when compared with other insect sequences) (Fig. 3D). Attempts to amplify further upstream or downstream coding sequences using different strategies always produced fragments containing stop codons in both directions. Orthology of the predicted protein with other insect TIM sequences was evident after phylogenetic analysis that included the paralog AcpTIM2 (Fig. 3C). However, as already described for AcpPER, although the sequenced fragment could be reliably aligned with other TIM sequences, the high divergence throughout the *A. pisum* sequence (Fig. 2) was reflected in its anomalous position in the phylogenetic tree and in the length of the branch leading to the *A. pisum* sequence (Fig. 3C). The presence of a somewhat modified NLS

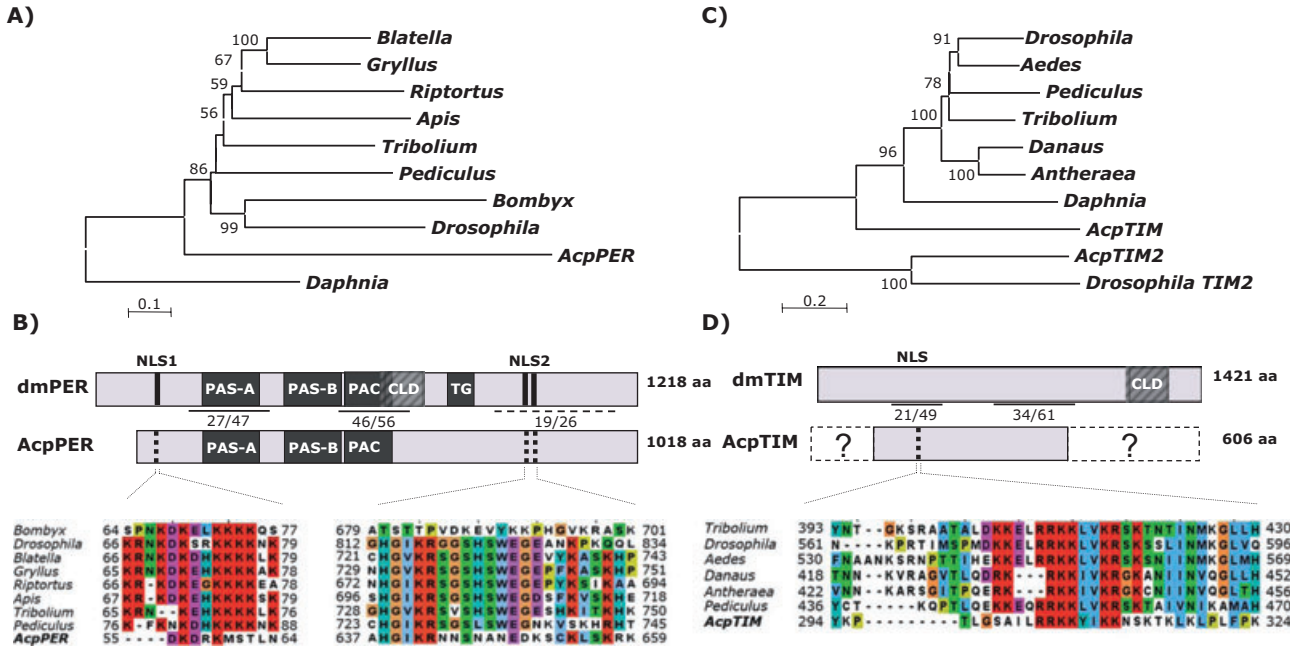

Figure 3. Phylogenetic analysis and structure comparisons of PER and TIM proteins in different species. (A and C) Phylogenetic relationships of insect PER and TIM proteins, respectively, inferred by neighbor joining on Poisson corrected distances. For PER, the *Daphnia* sequence was used as outgroup. In TIM analysis, the *Drosophila melanogaster* and *Acyrthosiphon pisum* sequences for the TIMEOUT paralog (here abbreviated as TIM2) were included as outgoups. Bootstrap values are shown for nodes only when higher than 50%. (B and D) Schematic representation of PER and TIM protein features from *D. melanogaster* and *A. pisum*. For *Drosophila*, isoform D of the TIM protein was used as reference. The numbers at the end of each diagram indicate protein size. For AcpTIM the filled box represents the sequenced portion of the protein while empty boxes with a question mark represent the hypothetical missing parts. Abbreviations: NLS, nuclear localization signal; CLD, cytoplasmic localization domain; PAS, PER-ARNT-SIM; TG, threonine-glycine region. Two PAS domains (PAS-A and PAS-B) followed by a PAC (PAS associated C-terminal) domain are found in PER protein. Black lines indicate the TIM and PER binding sites present in PER and TIM proteins, respectively, in *D. melanogaster* (Saez & Young, 1996). The dotted line indicates the CLK/CYC inhibition domain described in *Drosophila* (Chang & Reppert, 2003). Numbers below these lines represent percentage of identity between *D. melanogaster* and *A. pisum* and between *D. melanogaster* and *Pediculus humanus*, respectively. The alignments of different insect sequences corresponding to two PER and one TIM regions containing conserved NLS motifs are shown. Numbers on the left and right sides of each sequence in the alignment blocks indicate the relative positions of the first and last residues shown with respect to the N-terminus (position 1) for each sequence.

motif, that is highly conserved in other TIM sequences, was observed within a region relatively well conserved in other insect species but poorly conserved in *A. pisum* (Fig. 3D). In *Drosophila*, this motif has been shown to participate in the interaction with PER (Saez & Young, 1996).

Clock and cycle

Apart from minor point changes, predicted RefSeq models for *Clk* and *cyc* were coincidental with our cDNA derived sequences. Predicted proteins from both gene sequences contained the domains characteristic of bHLH-PAS transcription factors conserved in other insects and mammals (Hirayama & Sassone-Corsi, 2005). These included a 'basic helix-loop-helix' (bHLH) domain containing a well conserved NLS motif and PAS (A and B) and PAC domains (Fig. 4). The predicted AcpCLK protein was about 400 amino acids shorter than the corresponding *Drosophila* homolog having a size comparable to the homologous sequence in other insects (Fig. 4A; Rubin *et al.*, 2006). AcpCLK contained a C-terminal domain 60% similar to the Δ19 region found conserved in mouse and several insects (Chang *et al.*, 2003; Rubin *et al.*, 2006) that is apparently involved in the stabilization of the CLK-CYC heterodimer (Takahata *et al.*, 2000; Chang *et al.*, 2003). The polyglutamine-rich regions found in *Drosophila* are absent in AcpCLK and, hence, similar to the observations in *Antheraea pernyi* (Chang *et al.*, 2003).

The predicted AcpCYC protein was about 200 amino acids longer than the *Drosophila* homolog, as observed in other insects, and contained a C-terminal domain highly conserved in all insects studied except *Drosophila* which lacks this domain. This domain is highly similar to the transactivation domain described in the mouse CYC homolog BMAL (Takahata *et al.*, 2000; Chang *et al.*, 2003) (Fig. 4B). Although less dramatic than was observed for AcpPER and AcpTIM, both AcpCLK and AcpCYC seemed to accumulate numerous amino acid differences when comparing sequences from conserved aligned blocks with other insect sequences. These differences were reflected in the long branches leading to *A. pisum* sequences in the

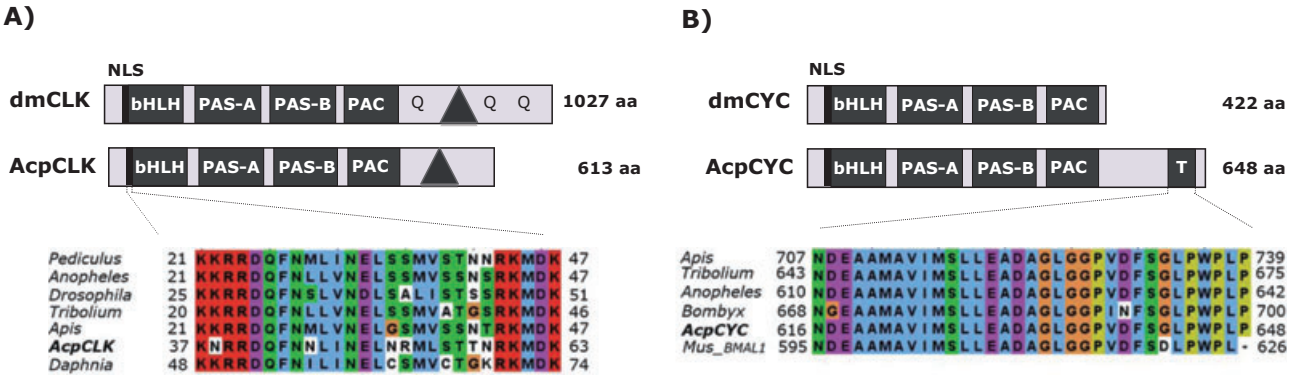

Figure 4. Schematic representation comparing the structure and some protein features in AcpCLK (A) and AcpCYC (B) with corresponding *Drosophila* proteins. The numbers at the end of each diagram indicate protein size. Abbreviations: NLS, nuclear localization signal; bHLH, basic helix-loop-helix domain; Q, polyglutamine-rich region; T, transactivation domain. PAS and PAC domains are as in Fig. 3. Triangle indicates a so-called Δ19 region (Chang *et al.*, 2003). Sequence alignments showing the degree of conservation of NLS and the transactivation domain in CLK and CYC, respectively, are shown. Numbers on the left and right sides of each sequence in the alignment blocks indicate the relative positions of the first and last residues shown with respect to the N-terminus (position 1) for each sequence.

phylogenetic trees obtained, and in the unexpected groupings of some sequences, especially for AcpCYC (Fig. S1). Although amino acid composition of AcpCLK and AcpCYC was not significantly different from other insects, relative rate tests showed that AcpCYC was evolving at accelerated rates in most comparisons while AcpCLK was accelerated in some of them (Table 1).

Vrille and Pdp1

SMART analysis of the *A. pisum* orthologs of these two basic zipper transcription factors revealed the presence of typical BRLZ (Basic-Region-Leucine-Zipper) domains in both of them with highly conserved DNA-binding domains (Fig. 5). As observed for other insects, both proteins were much shorter in *A. pisum* than in *Drosophila*. AcpVRI contained a glycine-serine rich stretch similar to *Drosophila* and a glutamine-rich domain which is absent from *Drosophila* VRI (Fig. 5A). AcpPDP1 contained a proline and acidic rich region (PAR) similar to *Drosophila* and other insects, but lacked the glutamine- and alanine-rich domains present in *Drosophila* (Cyran *et al.*, 2003). Interestingly, our cDNA sequence data, in agreement with the RefSeq prediction, revealed that AcpPDP1 appeared to lack seven amino acids from an extended basic domain including the highly conserved KKSRK amino acids that, in *Drosophila*, are necessary for proper DNA binding (Reddy *et al.*, 2000). However, a thorough analysis of sequencing traces derived from our directly sequenced PCRs and analysing additional recombinant plasmids containing cloned fragments from this gene showed a minority of transcripts including the missing seven amino acids (Fig. 5B). By looking at the genome sequence, it became clear that the seven amino acids corresponded to a 21 bp DNA sequence starting with a canonical GT donor site that was affected by alternative splicing – either incorporating this 21 bp or splicing it out as part of the adjoining intron. Finally, both AcpVRI and AcpPDP1 proteins had a homogeneous composition with other insect sequences and did not seem to be evolving at accelerated rates (Table 1).

Cryptochromes

Four different sequences that were highly similar to *Drosophila* CRY were found in the *A. pisum* genome by BLAST search. In order to ascertain orthology relationships, we built up a phylogenetic tree including representative sequences of the different insect and vertebrate *cryptochromes* including evolutionary related 6-4 DNA photolyases (Yuan *et al.*, 2007) (Fig. S2). The phylogenetic analysis revealed that both mammal-type and *Drosophila*-type *cryptochrome* genes were present in the *A. pisum* genome as well as a 6-4 DNA photolyase. Following Yuan *et al.* (2007) we designated the aphid cryptochromes as *AcpCry1* (*Drosophila*-type) and *AcpCry2* (mammal-type). We found *AcpCry2* is duplicated, with both versions differing by 3.5% of their total (18 out of 512 amino acids). Moreover, both versions of AcpCRY2 cDNAs were cloned and sequenced in our *A. pisum* strain (see Experimental procedures), revealing several polymorphisms when compared with the NCBI sequences (data not shown), which supported their paralogous character. SMART analysis showed that AcpCRYs contained typical flavin adenine dinucleotide (FAD) binding and photolyase domains (Fig. 6). Additionally, both AcpCRY2 paralogs, but not AcpCRY1, contained conserved domains that, in the mouse and zebrafish, are necessary for the CLK-CYC repressing function of mCRYs (Hirayama *et al.*, 2003; Rubin *et al.*, 2006). NLS and Coiled-coil motifs described in mammal-type CRYs (Hirayama *et al.*, 2003; Chaves *et al.*, 2006) were also conserved in AcpCRYs (Fig. 6).

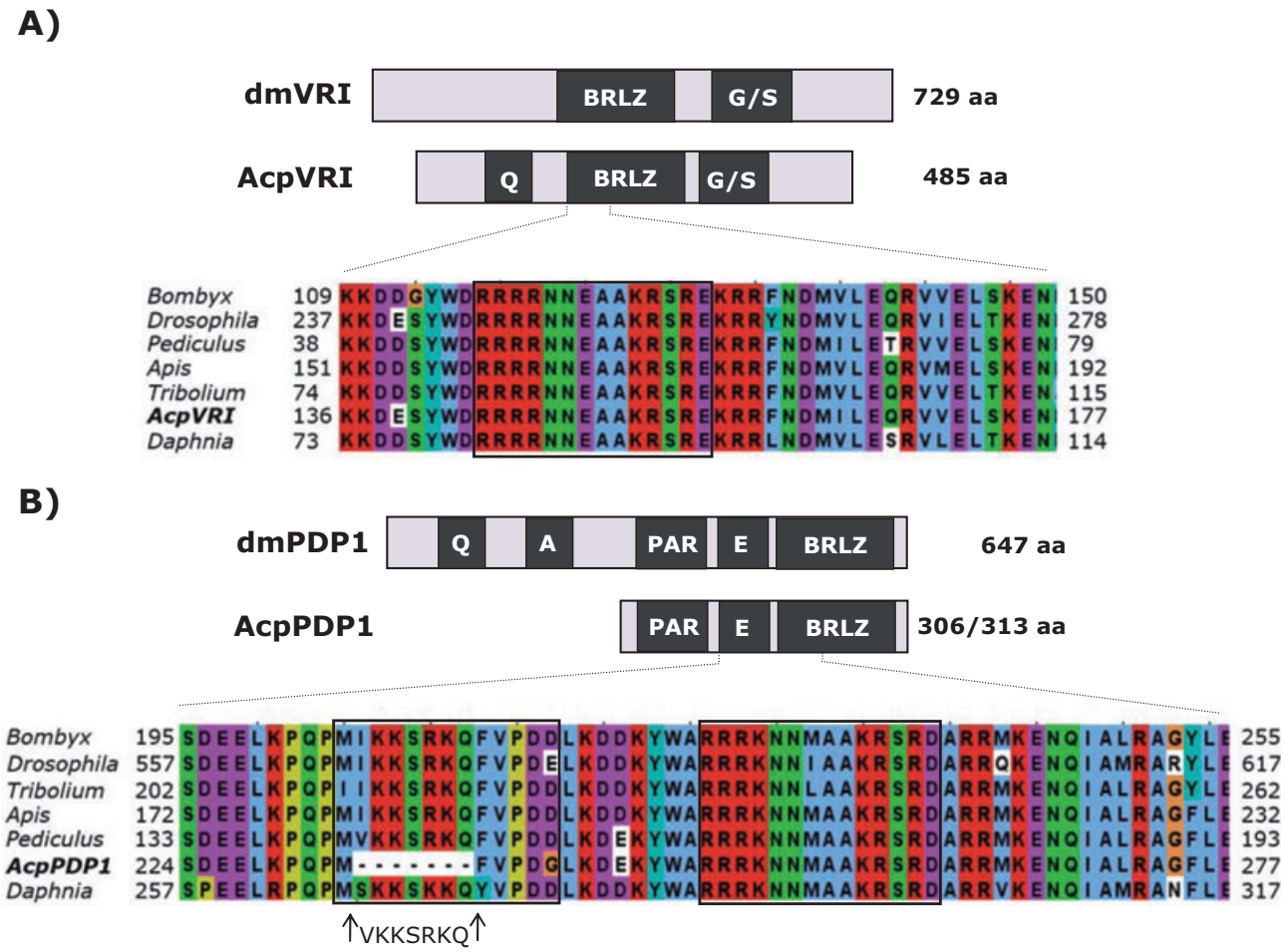

Figure 5. Schematic representation comparing the structure and some protein features in AcpVRI (A) and AcpPDP1 (B) with corresponding *Drosophila* proteins (isoforms A and D, respectively). The numbers at the end of each diagram indicate protein size. Abbreviations: BRLZ, basic region leucine zipper; G/S, glycine-serine rich stretch; Q, glutamine-rich domain; A, alanine-rich domain; PAR, proline and acidic rich domain; E, extended DNA-binding domain. Alignments containing partial BRLZ domains in different species for both genes are shown with boxes indicating well conserved DNA-binding domains. Numbers on the left and right sides of each sequence in the alignment blocks indicate the relative positions of the first and last residues shown with respect to the N-terminus (position 1) for each sequence. A 'VKKSRKQ' motif between arrows indicates a second AcpPDP1 splicing variant found containing this motif apart from the most abundant variant lacking these seven amino acids (included in the alignment).

Interestingly, both AcpCRY2 genes showed significant accelerated rates of change in most comparisons (Table 1; Fig. 2).

Expression analysis of core-clock genes in Acyrthosiphon pisum

Temporal expression patterns of putative core-clock genes analysed through quantitative real-time PCR (see Experimental procedures) revealed weak to moderate oscillations in mRNA abundance for some genes in aphid heads along the day–night cycle, especially in aphids reared under long-day (LD) conditions (Fig. 7). In general, the oscillating patterns observed in mRNA levels in heads of aphids reared under short-day conditions (SD) were similar to those observed in aphids reared under long-day conditions (LD). However, for some genes, a significantly elevated expression in SD aphids was observed (Fig. 7). This was the case for *AcpPer*, *AcpPdp1*, and *AcpCry2* (Fig. 7) and also *AcpCry1* (not shown). No significant oscillation along the day–night cycle was observed for *AcpClk* and *AcpCry1*. Conversely, a significant oscillation was observed in *AcpCyc* under LD, with a peak of expression at the beginning of the light phase (Fig. 7, $P = 0.04$). *AcpPer* also significantly oscillated in LD aphids with the maximum expression at the end of the light phase (Fig. 7, $P = 0.0002$). As for other genes, oscillation of *AcpPer* under SD conditions was not significant, likely attributable to high variability at some time points. *AcpCry2* oscillated both in LD and SD and, interestingly, showed the same temporal pattern of expression as *AcpPer* under both photoperiod conditions (Fig. 7, $P = 0.05$ and $P = 0.04$,

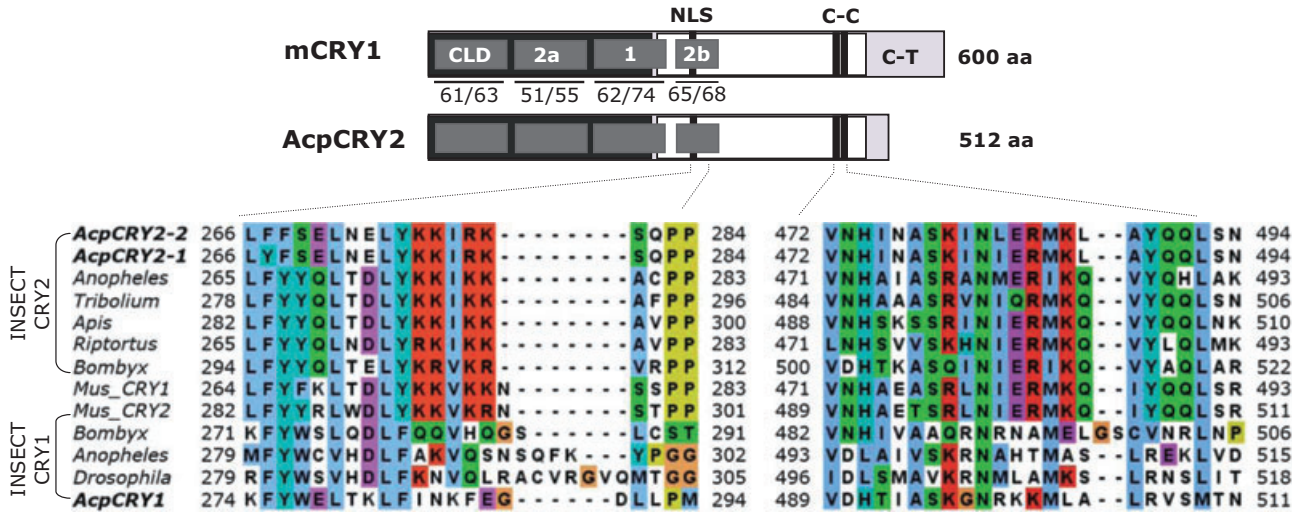

Figure 6. Schematic representation comparing the structure and some protein features in the mouse mCRY1 with its aphid ortholog AcpCRY2. The numbers at the end of each diagram indicate protein size. Dark grey and white backgrounds demarcate FAD- binding and photolyase domains, respectively. Three regions (RD-2a, RD-1 and RD-2b) necessary for the CLK-BMAL repressing function are shown as 2a, 1 and 2b, respectively (Hirayama *et al*., 2003). Percentage identity between *Mus musculus* and *A. pisum* and between *M. musculus* and *Riptortus pedestris* (Ikeno *et al*., 2008), respectively, are indicated for these three regions and for a CLD domain. The C-terminal region (C-T) is indicated. All mouse features conserved in the aphid are indicated on the aphid molecule by similar boxes. Two multiple sequence alignments corresponding to the mouse RD-2b region containing an NLS motif and to a Coiled-coil domain (C-C), both necessary for nuclear entry (Chaves *et al*., 2006) are shown. The alignments also include AcpCRY1 and other insect sequences representative of the *Drosophila*-type *cryptochrome* (Insect CRY1) showing the lack of the above domains in this paralog. Numbers on the left and right sides of each sequence in the alignment blocks indicate the relative positions of the first and last residues shown with respect to the N-terminus (position 1) for each sequence.

respectively). *AcpVri* and *AcpPdp1* both oscillated in LD aphids with a pattern similar to that observed for *AcpCyc*, having maximum expression at the beginning of the light phase and the minimum at the end of the light phase (Fig. 7, $P = 0.05$ and $P = 0.005$, respectively). *AcpVri* also significantly oscillated in SD aphids ($P = 0.008$). Finally, expression of *AcpTim* was also rhythmic in LD, with a peak at the beginning of the dark phase (as observed for *AcpPer*). However, difficulties experienced in the amplification of *AcpTim* cDNAs (including the frequent presence of dimers and the occasional amplification of fragments of unexpected size) suggest our results for *AcpTim* should be treated with caution until additional data becomes available. These problems led us to exclude from the analysis *AcpTim* results from SD aphids.

Discussion

In the present study we have described the gene repertoire of the putative circadian clock of the aphid *Acyrthosiphon pisum* based on the analysis of its recently available genome. We have also analysed the structure, evolution and the expression of genes at the core of the clockwork. Although some individual clock genes have been described in other species, this is the first hemimetabolous insect in which the full core clockwork gene set has been investigated and it will therefore contribute to a better understanding of the functioning and evolution of the circadian clock in insects.

We have shown the presence in *A. pisum* of orthologous sequences of the full gene set constituting the core of the two feedback loops described in the *Drosophila* circadian clock, as well as most of the ancillary genes that control the stability/degradation of core clock proteins (Fig. 1; Table 1). Notably, the *A. pisum* genome contains a mammalian-type *cryptochrome* (*AcpCry2*), confirming its presence in hemimetabola (see Ikeno *et al*., 2008) and suggesting it has been lost in drosophilids (Rubin *et al*., 2006; Yuan *et al*., 2007; Zhu *et al*., 2008). Moreover, the presence of two copies of this gene in the aphid is similar to the situation in many vertebrates, but this is the first time it is reported in an insect, which points to an additional relevance for this aphid gene. In addition, contrary to the situation in *T. castaneum* and *A. mellifera* but similar to Lepidoptera, *A. pisum* also contains a copy of a *Drosophila*-type *cryptochrome* (*AcpCry1*), which would place the aphid clockwork closer to the hypothetical ancestral clock model proposed by Yuan *et al.* (2007) as in the butterfly (see Introduction). Under this model, AcpCRY1 would work primarily as a circadian photoreceptor whereas AcpCRY2 would be an essential component of the core clockwork acting as a major transcriptional repressor of the clock alone, or in conjunction with AcpPER (see below). The presence of almost the entire

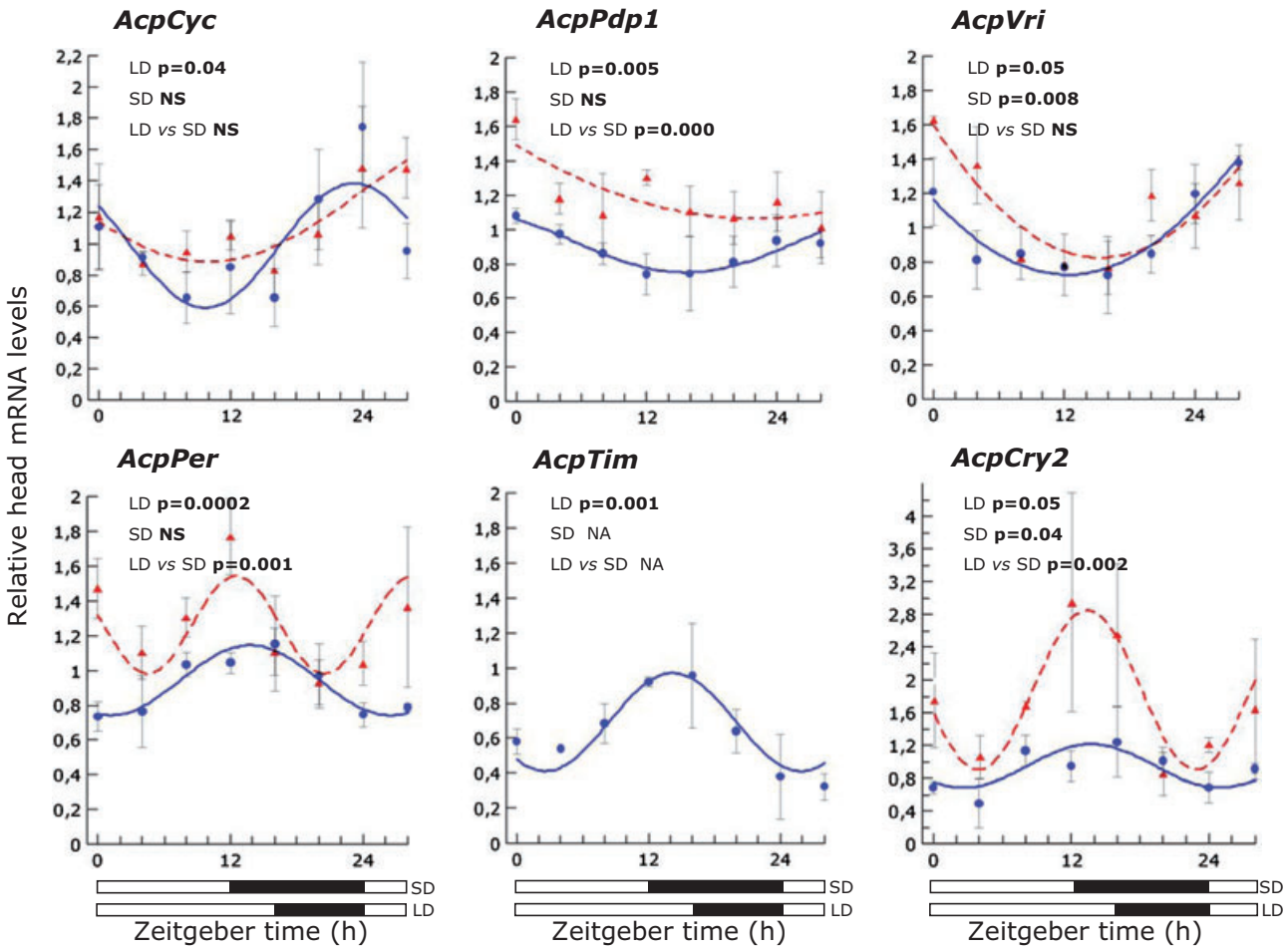

Figure 7. Temporal expression patterns of core clock genes in *Acyrthosiphon pisum*. The plots show the average (± SE) relative head mRNA levels for each time point, represented as filled circles and filled triangles for LD and SD regimes, respectively (see Experimental procedures). Continuous lines and dotted lines represent sine wave curves fitted to 4 h interval data (see Experimental procedures). Open bars at the bottom of the plot indicate light and black bars indicate dark during the two illumination regimes used. For each gene *P*-values obtained after the zero-amplitude test for cosinor analysis are shown when significant. One way ANOVA was performed in order to check for significant differences in relative head mRNA levels between the two photoperiods investigated. *P*-values are indicated for those genes with significant differences between LD and SD photoperiods. NS, non significant; NA, not applicable.

clockwork gene set in the aphid genome (with the exception of a *jetlag* ortholog, see below) might suggest a high conservation of the clock mechanism. However, a detailed analysis of individual genes shows that the aphid system is somewhat unique. First, the *A. pisum* orthologs of the four transcription factors at the core of the *Drosophila per*/*tim* feedback loop (AcpPER, AcpTIM, AcpCYC and AcpCLK) are present in the aphid but are highly divergent and are evolving faster than corresponding proteins in other insects. This is especially true for AcpPER and AcpTIM (Table 1; Fig. 2). Similarly, the two copies of AcpCRY2 are also evolving at accelerated rates. If, as has been proposed for other insects (Rubin *et al.*, 2006; Yuan *et al.*, 2007; Zhu *et al.*, 2008), AcpCRY2 does have a central role in the aphid clock, then all genes at the core of the putative *per*/*tim* loop in the aphid oscillator are changing rapidly with the exception of AcpCRY1.

The possibility that the observed accelerated rates of change in the genes of the *per*/*tim* loop is reflecting a different clockwork system in hemimetabola can be ruled out as sequence comparisons always included at least a second hemimetabolous insect (usually the human lice *P. humanus*, whose genome is also currently being sequenced) not showing these high rates of sequence change. This was most evident after the analysis of AcpPER which, in addition to the *P. humanus* sequence, also included sequences from distant hemimetabolous representatives (*Blatella* and *Gryllus*) and from the hemipteroid *Riptortus pedestris* (Fig. 3A; Ikeno *et al.*, 2008) which are more similar to *Drosophila* and other holometabola sequences than the *A. pisum* sequence. Similarly, the sequenced portion of AcpTIM was found to be more divergent than the *Daphnia* ortholog when compared to other insects (Fig. 3C). Therefore the high divergence

observed for core clock genes would be characteristic of aphid sequences and not the result of a holo- vs. hemimetabolous differential clock evolution. Although it is possible that the genes at the core of the *per*/*tim* feedback loop in the aphid are each evolving independently to adapt to new aphid-specific functions, it seems more reasonable to think that the observed high rates of change in all these genes are reflecting a co-ordinated evolution to adapt the clock to the (as yet unknown) aphid-specific needs. It is worth to mention in this context the unexpected phylogenetic position reported for the PAS dimerization domain in the *Musca domestica* PER protein (Piccin *et al*., 2000) which, along with a functional assay, led the authors to hypothesize an intermolecular coevolution between PER and TIM. Contrary to observations in core clock genes, the other (ancillary) genes identified in the aphid's genome, putatively involved in the stability/degradation of the aphid core clock proteins, all seemed to be evolving according to the pace in other insect species, which agrees with these genes having additional non-clock related functions (Vallone *et al*., 2007). Moreover, two of the three genes participating at the putative aphid second feedback loop (genes *AcpVri* and *AcpPdp1* controlling gene *Clock* expression, see Fig. 1) did not show any evidence of accelerated rates of change (Table 1; Fig. 2), which agrees with their orthologs being, at least in *Drosophila*, transcription factors controlling the expression of other genes in addition to *Clock* (see reports for both genes at http://flybase.org/). Therefore the distinctive pattern observed for genes at the core of the aphid circadian clock reflects a process of adaptation of the aphid clockwork rather than a global process affecting all aphid genes.

More studies are obviously needed to know the force(s) behind the fast evolution of aphid core clock genes. However, our present results on sequence analysis and expression data on particular genes may help to shed some light on the issue and constitute the starting point for further research. For example, in the case of the aphid PER protein, the high divergence detected at the amino acid level also affects two NLS motifs (highly conserved in all other insects) that have been proposed to be involved in the nuclear import of the protein, which are absent or highly modified (Fig. 3B). As nuclear import is essential for the transcription inhibition role attributed to PER, it is likely that AcpPER on its own lacks this function which, along with the high substitution rate detected, might be indicative of a modified functionality of this gene. However, it is also possible that the second NLS, by conserving some basic residues in *A. pisum* (Fig. 3B), is still functional, allowing independent entry of AcpPER into the nucleus to exert its repressing role on the putative AcpCLK-AcpCYC dimers. Finally, we cannot disregard that AcpPER may enter the nucleus with the help of another protein. Both AcpTIM and AcpCRY2 would be the candidates for this role as their orthologs in *Drosophila* and the mouse, respectively, have been shown to stabilize and help PER nuclear entry (Saez & Young, 1996; Hirayama *et al*., 2003; Sakakida *et al*., 2005). Our data do not show if these interactions take place among these proteins in the aphid. However, the high divergence of putative domains that, in *Drosophila*, are involved in the binding of PER and TIM, and in the interaction with the CLK-CYC dimer (Fig. 3B, D), suggests that these interactions may not occur in the aphid. Since AcpCRY2 domains, that participate in nuclear entry and repression of CLK-CYC activity in mammalian-type Cryptochrome (Hirayama *et al*., 2003), are highly conserved (Fig. 6), this strongly supports a conserved repressing function of this aphid gene. Whether such repression requires AcpPER remains to be elucidated. With respect to the aphid orthologous of *Drosophila* PER's partner (AcpTIM), the fact that it is even more divergent than AcpPER (its definitive size pending future research) supports the thesis that it has evolved a modified functionality, either within or outside the clock machinery. In this context, the absence of a *jetlag* homolog in the *A. pisum* genome provides support, albeit indirect, for a modified functionality (or even a loss of function) of AcpTIM. JETLAG (JET) is an F-box protein of the SCF E3 ubiquitin ligase family that, in *Drosophila*, has been shown to be essential for synchronising the clock with the day–night cycles by promoting the degradation of TIM in response to light (Koh *et al*., 2006). It has also been demonstrated that CRY1 is necessary in the process (as the light photoreceptor transmitting photic information) and that both proteins must physically interact with TIM (Koh *et al*., 2006). We investigated the presence of *jet* in different insect species with full genome sequence available and found that it was only absent in species in which either *cry1* or *tim* orthologs were missing (e.g. *T. castaneum*, *A. mellifera* and *P. humanus*; own unpublished results). Therefore, the absence of JET in *A. pisum* may indicate that, if aphid clock cells have an intrinsic light entrainment mechanism, it may not rely on AcpTIM. We do not discount the possibility that the accelerated evolution detected in *A. pisum per* and *tim* genes may be the result of selection pressure on functions of these genes not related with the circadian clock. In this respect, both *per* and *tim* genes in *Drosophila* have been involved in different species-specific functions directly related with reproductive success, including courtship and mating behaviour (Kyriacou & Hall, 1980, 1986; Sakai & Ishida, 2001; Tauber *et al*., 2003; Beaver & Giebultowicz, 2004) and gamete production (Beaver *et al*., 2002, 2003; Kotwica *et al*., 2009).

Analysis of the expression of putative aphid clock genes should provide clues on their roles (if any) in the clockwork. In this respect, the difference in the number of ESTs present in public databases between core clock genes

and genes involved in stability/degradation of core products (Table 1) is most notable. Attending to those data, the expression of core clock genes in the aphid must be rather limited compared with the expression of ancillary genes which are involved in various other processes apart from their putative roles in the circadian clock. Our experimental results on the expression of core clock genes draws a preliminary picture of how the circadian clock works in *A. pisum*. First, significant rhythm in transcript levels for some genes shown in Fig. 7 shows their regulation in response to the zeitgeber (the light–dark cycles) and would suggest these genes are also part of the *A. pisum* endogenous circadian oscillator. Second, the coincident phasing of peak *AcpPer*, *AcpTim* and *AcpCry2* transcripts (our assay did not distinguish between the two versions of the latter gene), which all peak at the end of the day, points to the coordinate regulation of these genes. Whether their expression is activated by AcpCLK-AcpCYC dimers binding to common elements in their promoter regions will need further investigation, including the search for E-boxes in the promoter regions of these genes. In *Drosophila* both *per* and *tim* mRNAs peak synchronously at dusk (within the dark phase; Williams & Sehgal, 2001). The same pattern of expression has been observed in butterflies for orthologs of these genes (Zhu *et al*., 2008) and in *A. mellifera* for *amPer* and *amCry2* genes (a *timeless* ortholog is absent in the honey bee) (Rubin *et al*., 2006). The pattern observed for the orthologs of these three genes in *A. pisum* was similar to the pattern described in these organisms but somewhat advanced: in the aphid these genes were shown to peak at the end of the day rather than at the beginning of the dark phase (Fig. 7). Third, similar to *Apis mellifera* (Rubin *et al*., 2006) *AcpCyc*, but not *AcpClk*, showed a significant oscillation with maximum expression at dawn, in anti-phase with the expression of *AcpPer*, *AcpTim* and *AcpCry2* (Fig. 7). This, along with the high conservation in *A. pisum* of a transactivation domain present at the C-terminal region of CYC proteins in all non-drosophilid insect species but also in the mammalian ortholog BMAL (Fig. 4B), places the aphid clock closer to the mammal than to the *Drosophila* system. It is worth remembering that in the mouse *mBmal*, and not *mClk*, oscillates which is contrary to the situation in *Drosophila* (Stanewsky, 2003). Fourth, the synchronous oscillation of both *AcpPdp1* and *AcpVri* in antiphase with *AcpPer*, *AcpTim* and *AcpCry2* is different from *Drosophila* where their expression is also controlled by the CLK-CYC dimer (Cyran *et al*., 2003). This fact and the previous discussion on *AcpCyc* and *AcpClk* expression suggest that the second loop works differently in the aphid circadian clock and probably in other insects (Fig. 1). Unfortunately, only two studies on the expression of *vrille* (Ikeno *et al*., 2008) and *Pdp1* (Dolezel *et al*., 2008) orthologs in other insects are available, which prevents us from making comparisons in order to make a general statement on the role/s, if any, of these two genes in the circadian clock in non-drosophilid insects. Fifth, the lack of any evident oscillation of *AcpCry1* is similar to observations in *A. pernyi* (Zhu *et al*., 2008) compatible with a mere role in photoreception rather than as a core pacemaker. Sixth, according to our results modifications of the photoperiod regime do not appear to substantially alter the oscillating pattern of expression of core clock genes. However, a significant increase in the expression of some of them was observed in aphids reared under short days suggesting a link between the control of circadian rhythm and photoperiod controlled processes. It would be tempting to propose that the aphid clockwork has become biased towards the control of seasonal (i.e. photoperiodic), rather than daily, responses. Most relevant in this context, several reports have shown the involvement of orthologs of clock genes *timeless* and *period* in photoperiodic responses (Goto & Denlinger, 2002; Pavelka *et al*., 2003; Shimada, 2005; Goto *et al*., 2006; Mathias *et al*., 2007; Stehlik *et al*., 2008), supporting the view that photoperiodism in insects relies on the circadian clock machinery. It has also been shown that natural selection has shaped genetic variation at the *period* and *timeless* loci in *Drosophila* producing latitudinal clines which are usually interpreted as evidence of adaptation to seasonal conditions at different latitudes (Costa *et al*., 1992; Sandrelli *et al*., 2007; Tauber *et al*., 2007; reviewed in Kyriacou *et al*., 2007). However, as discussed by Bradshaw & Holzapfel (2007a,b), it is possible that circadian genes are only incidentally involved in photoperiodism and both circadian and seasonal rhythms would rely on independent gene circuits.

Finally, we have shown that *A. pisum* has the set of genes necessary for a circadian clock to function. Moreover, the expression observed for some genes suggest a clockwork system closer to the hypothetical ancestral system in insects rather than to the *Drosophila* system (Ikeno *et al*., 2008). In this system, AcpCRY2 (alone or together with AcpPER) would have an essential role as repressor of a putative AcpCLK-AcpCYC-induced transcription. It is likely that AcpCRY1 works in this system as a photoreceptor synchronizing in some unknown way the clock with the day–night cycles. It is also likely that the second feedback loop described in *Drosophila* (Fig. 1) is different in *A. pisum* and that a loop centered on *AcpCyc* instead of *AcpClk* works in aphids similar to vertebrates. However, despite having all the necessary elements and having shown that expression of some genes are compatible with a circadian clock working in the aphid, in order to demonstrate that the studied genes are bona fide components of the circadian mechanism in the pea aphid, it is necessary to show that their mRNAs continue to oscillate in constant darkness (DD, Bell-Pedersen *et al*., 2005). Although in the present report we have put forward some

hypotheses, there are some additional questions that need immediate clarification. For instance, why are some clock genes, especially AcpPER and AcpTIM, evolving so fast? Has AcpTIM (or AcpPER) been recruited for a different function? Do any of the aphid genes participate in the photoperiodic (seasonal) rhythmicity? We believe that full length sequencing of *AcpTim* and more thorough analysis of the expression of core clock genes (including *in situ* hybridizations) are urgently needed to address these questions. Although much work remains, our analysis provides the grounds for subsequent studies aimed at understanding how the clock works in aphids and at elucidating the possible role of circadian clock elements in photoperiodism.

Experimental procedures

Identification of clock gene homologs in Acyrthosiphon pisum

The pea aphid genome assembly Acyr 1.0 (http://www.ncbi.nlm.nih.gov/projects/genome/guide/aphid/) was used to search for homologs of *Drosophila melanogaster* genes known to have a role in the circadian clock in this organism (Table 1; Fig. 1 for a comprehensive list of the genes). Different BLAST searches (Altschul *et al*., 1997) were done on *Drosophila* amino acid sequences, using both NCBI and AphidBase (http://www.aphidbase.com/aphidbase/) resources. When alternative protein products were available for a *Drosophila* gene, the longest one was chosen. Initially, the searches were carried out against *A. pisum* RNA (TBLASTN) and protein (BLASTP) NCBI's RefSeq databases, followed by queries against the genome scaffolds available in AphidBase. Additionally we conducted our searches across *ab initio* and other non-RefSeq prediction databases included within NCBI and AphidBase blasting options. TBLASTN searches were also carried out against unassembled genome sequences. Queries were also implemented against pea aphid EST databases to look for transcript support of gene models. Finally, searches against the WGS traces database available at NCBI were also performed in some instances.

Sequencing of cDNAs from putative Acyrthosiphon pisum *clock genes*

Genomic sequences and mRNA predictions of *A. pisum* homologs of *Drosophila per, tim, Clk, cyc, Pdp1, vri and cry* (including both *Drosophila* and mammal *cryptochrome* types) were aligned, which allowed us to design primers for PCR amplification based on exon sequences (Table S1). In addition to primers based on translated portions along the genes, we also designed several primers based on sequences located at various distances upstream of the initiation methionine and downstream of the stop codon. Our primer design allowed us to amplify each cDNA in two or more overlapping fragments that, once assembled using the Staden package (Staden *et al*., 1998), spanned the whole coding sequence along with partial 5′ and 3′ untranslated regions. Additionally, 3′ RACE, using forward gene specific primers and oligo-dT-T7 primers, were carried out to amplify the 3′UTRs. cDNA templates for PCR amplification were reverse transcribed (High Capacity cDNA reverse transcription kit, Applied Biosystems, Foster City, CA, USA) from total RNA obtained from whole aphids of our YR2 strain (Ultraspec-II RNA isolation system, Biotecx). PCR amplified products (5 PRIME Taq DNA polymerase) were usually sequenced directly after purification (High Pure PCR Product Purification Kit, Roche, Indianapolis, IN, USA) using previously designed primers. In some instances cloning of the PCR amplified fragments was necessary. The pGEM-T easy Vector System (Promega, Madison, WI, USA) was used for this purpose. Because of recurrent amplification problems with the different PCR strategies used, only partial sequences of *timeless* and *vrille* aphid orthologs have been characterized (about 50% of the expected coding sequence for both genes).

Our novel cDNA sequences allowed us to check the gene models or predictions available for the 8 sequenced genes. For the rest of the putative clock genes identified (Table 1) the gene models were checked by comparing the predicted amino acid sequences with the orthologous sequences in other insects (see below) and with the ESTs available for some of the genes.

Phylogenetic analyses

To confirm the orthology of the putative clock genes identified in the pea aphid genome, we carried out phylogenetic analyses that included orthologous sequences from other organisms (Table S2). Sequences from insect species with sequenced genome available, representative of different insect orders, were used besides sequences of *A. pisum* and *D. melanogaster* (e.g. *Bombyx mori*, *A. mellifera* and *T. castaneum*). We also included in the analysis the sequences from *P. humanus*, whose genome is currently being sequenced (http://phumanus.vectorbase.org/). Other insect sequences were also used for the phylogenetic analysis of some of the genes when available (Table S2). The orthologous sequences of the water flea *Daphnia pulex* were generally used as outgroup sequences. The orthologs of the clock genes present in representative vertebrate genomes were also included in most of the phylogenetic analyses. For some genes, the sequences of certain paralogous genes were also included in the phylogenetic analyses (e.g. *Timeout* and *Tango*, paralogs of *timeless* and *Clock* and *cycle*, respectively). Other paralogs included in the analysis but not shown in Table S2 are *B56, 6-4 photolyase, Casein kinase I* α and *Gilgamesh,* and *Archipelago*, paralogs of *widerborst*, *cryptochromes*, *double-time* and *supernumerary limbs*, respectively. To ease the display, most trees shown in the different figures only include insect and *Daphnia* orthologs.

Amino acid sequences were initially aligned using CLUSTALX v1.81 (Thompson *et al*., 1997) in combination with MEGA v4.0 (Tamura *et al*., 2007), and subsequently revised manually. The alignment of the pea aphid protein predictions to other insect orthologs was used in combination with *A. pisum* gene models and ESTs (when available) to check the predicted gene structure and to identify exon and intron boundaries in the genomic sequence. Usually, the presence of amino acid stretches in *A. pisum* protein predictions not present in the rest of insect proteins or the lack of regions conserved in the other species were used as indications for likely wrong predictions in the gene models. Phylogenetic analyses were always carried out using only those reliably aligned regions of the genes (Table 1). Overall and pairwise distances among the sequences were calculated using MEGA. The homogeneity of aminoacid composition of sequences

was evaluated using TREE-PUZZLE v5.2 (Schmidt *et al*., 2002). The program RRTree (Robinson-Rechavi & Huchon, 2000) was chosen to test for homogeneity in rates of sequence evolution. Phylogenetic trees were constructed with the neighbor-joining algorithm as implemented in MEGA, using Poisson corrected distances. Finally, node support was evaluated using 1000 bootstrap replicates (Felsenstein, 1985).

Putative functional domains and motifs on the pea aphid clock genes

Several functional domains and motifs are well described for the clock proteins in *Drosophila*, and other insects as well as in some vertebrates (Saez & Young, 1996; Chang & Reppert, 2003; Chang *et al*., 2003; Chaves *et al*., 2006; Hirayama *et al*., 2003; Hirayama & Sassone-Corsi, 2005; Sakakida *et al*., 2005; Rubin *et al*., 2006, among others). The descriptions of relevant protein regions in the literature were used to demarcate these sequences on the putative clock genes in *A. pisum*. The SMART server (Schultz *et al*., 1998) was also used to find some motifs and domains. Partial alignments shown in different figures were obtained from full alignment clustal files and manipulated using Jalview (Clamp *et al*., 2004).

Analysis of expression

For expression analysis, the *A. pisum* strain LSR1 used for genome sequencing was used. Aphids were kept on broad bean plants as a parthenogenetic reproducing clone under long day conditions (16L:8D, 18 °C). Aphids were maintained at low density conditions (up to 6 aphids per plant) over three generations in order to avoid the formation of winged forms. Aphids from two photoperiod conditions were used for expression analysis. LD aphids were obtained from the usual long day conditions (16L:8D). SD aphids were obtained from the second generation (G1) of aphids reared under short day conditions (12L:12D) that usually produce the sexual response (Cortes *et al*., 2008). Adult aphids from both conditions were collected and frozen in liquid nitrogen at 4 h intervals (starting at ZT0, when lights went on) along the day–night cycle. Aphids were kept at −80 °C until RNA extraction. Visual inspection of the SD descendants (G2) confirmed that induction of sexuality had occurred. Three different groups of 10 aphids were pooled from each time point and photoperiod condition. Heads were cut and separated from the aphid bodies on a Petri dish on a dry ice layer. Total RNA was extracted from heads and bodies separately using the Ultraspec-II RNA isolation system (Biotecx Laboratories, Houston, TX). RNA was treated for DNase contamination using the DNA-free™ Kit (Applied Biosystems). One microgram of total RNA was reverse transcribed with random primers using the High Capacity cDNA reverse transcription kit (Applied Biosystems).

Temporal expression of putative clock genes *AcpPer, AcpTim, AcpClk, AcpCyc, AcpPdp1, AcpVri, AcpCry1* and *AcpCry2* in aphid heads and bodies was analyzed by real-time quantitative PCR using the StepOnePlus™ Real-Time PCR System with the Power SYBR® Green PCR Master Mix (Applied Biosystems). Specific primers were designed in order to ensure amplification of discrete bands with no primer-dimers. (Table S1). Expression along the day–night cycle for both photoperiod regimes was analysed in duplicate. The *RpL7* gene was used as endogenous control of constitutive expression (Nakabachi *et al*., 2005). Relative expression was calculated using the Pfaffl method (Pfaffl, 2001) on values normalized to the arithmetic mean of Ct values across all samples for each gene within a run. MATLAB (The MathWorks, Inc) was used to perform Cosinor analysis (Nelson *et al*., 1979) to test for rhythmicity on relative expression values obtained as described above. Rhythm parameters were estimated by a least-squares fitting of a cosine function and statistical significance was evaluated by performing zero amplitude tests. Sine wave curves were fitted to 4-h interval data by using XL-fit software (IDBS).

Acknowledgements

We thank BCM-HGCS for making the *Acyrthosiphon pisum* genome publicly available. We thank Dr Hugh M. Robertson for his help in identifying aphid *per* and *tim* homologs. This work was supported by project CGL2007-67392/BOS from the Spanish MEC. TC received a PhD grant from the FPI program, MEC, Spain. The facilities at SCSIE (Universitat de València) were used for sequencing and RT-Q-PCR and for plant growth for aphid rearing.

References

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. *et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389–3402.

Beaver, L.M. and Giebultowicz, J.M. (2004) Regulation of copulation duration by *period* and *timeless* in *Drosophila melanogaster*. *Curr Biol* **14**: 1492–1497.

Beaver, L.M., Gvakharia, B.O., Vollintine, T.S., Hege, D.M., Stanewsky, R. and Giebultowicz, J.M. (2002) Loss of circadian clock function decreases reproductive fitness in males of *Drosophila melanogaster*. *Proc Natl Acad Sci USA* **99**: 2134–2139.

Beaver, L.M., Rush, B.L., Gvakharia, B.O. and Giebultowicz, J.M. (2003) Noncircadian regulation and function of clock genes *period* and *timeless* in oogenesis of *Drosophila melanogaster*. *J Biol Rhythms* **18**: 463–472.

Bell-Pedersen, D., Cassone, V.M., Earnest, D.J., Golden, S.S., Hardin, P.E., Thomas, T.L. *et al.* (2005) Circadian rhythms from multiple oscillators: lessons from diverse organisms. *Nat Rev Genet* **6**: 544–556.

Bradshaw, W.E. and Holzapfel, C.M. (2007a) Evolution of animal photoperiodism. *Annu Rev Ecol Evol Syst* **38**: 1–25.

Bradshaw, W.E. and Holzapfel, C.M. (2007b) Tantalizing timeless. *Science* **316**: 1851–1852.

Bradshaw, W.E., Quebodeaux, M.C. and Holzapfel, C.M. (2003) The contribution of an hourglass timer to the evolution of photoperiodic response in the pitcher-plant mosquito, *Wyeomyia smithii*. *Evology* **57**: 2342–2349.

Chang, D.C. and Reppert, S.M. (2003) A novel C-terminal domain of *Drosophila* PERIOD inhibits dCLOCK:CYCLE-mediated transcription. *Curr Biol* **13**: 758–762.

Chang, D.C., McWatters, H.G., Williams, J.A., Gotter, A.L., Levine, J.D. and Reppert, S.M. (2003) Constructing a

feedback loop with circadian clock molecules from the silkmoth, *Antheraea pernyi*. *J Biol Chem* **278**: 38149–38158.
Chaves, I., Yagita, K., Barnhoorn, S., Okamura, H., van der Horst, G.T.J. and Tamanini, F. (2006) Functional evolution of the photolyase/cryptochrome protein family: importance of the C terminus of mammalian CRY1 for circadian core oscillator performance. *Mol Cell Biol* **26**: 1743–1753.
Clamp, M., Cuff, J., Searle, S.M. and Barton, G.J. (2004) The Jalview Java alignment. *Bioinformatics* **20**: 426–427.
Collins, B., Mazzoni, E.O., Stanewsky, R. and Blau, J. (2006) *Drosophila* CRYPTOCHROME is a circadian transcriptional repressor. *Curr Biol* **16**: 441–449.
Cortes, T., Tagu, D., Simon, J.C., Moya, A. and Martinez-Torres, D. (2008) Sex versus parthenogenesis: a transcriptomic approach of photoperiod response in the model aphid *Acyrthosiphon pisum* (Hemiptera: Aphididae). *Gene* **408**: 146–156.
Costa, R., Peixoto, A.A., Barbujani, G. and Kyriacou, C.P. (1992) A latitudinal cline in a *Drosophila* clock gene. *Proc Biol Sci* **250**: 43–49.
Cyran, S.A., Buchsbaum, A.M., Reddy, K.L., Lin, M.C., Glossop, N.R., Hardin, P.E. *et al.* (2003) vrille, Pdp1, and dClock form a second feedback loop in the *Drosophila* circadian clock. *Cell* **112**: 329–341.
Dolezel, D., Zdechovanova, L., Sauman, I. and Hodkova, M. (2008) Endocrine-dependent expression of circadian clock genes in insects. *Cell Mol Life Sci* **65**: 964–969.
Eisenbach, J. and Mittler, T.E. (1980) An aphid circadian rhythm: factors affecting the release of sex pheromone by oviparae of the greenbug, *Schizaphis graminum*. *J Insect Physiol* **26**: 511–515.
Elsik, C., Mackey, A., Reese, J., Milshina, N., Roos, D. and Weinstock, G. (2007) Creating a honey bee consensus gene set. *Genome Biol* **8**: R13.
Emerson, K.J., Dake, S.J., Bradshaw, W.E. and Holzapfel, C.M. (2009) Evolution of photoperiodic time measurement is independent of the circadian clock in the pitcher-plant mosquito, Wyemoyia smithii. *J Comp Physiol A Neuroethol Sens Neural Behav Physiol* **195**: 385–391.
Felsenstein, J. (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evology* **39**: 783–791.
Goto, S.G. and Denlinger, D.L. (2002) Short-day and long-day expression patterns of genes involved in the flesh fly clock mechanism: *period, timeless, cycle* and cryptochrome. *J Insect Physiol* **48**: 803–816.
Goto, S.G., Han, B. and Denlinger, D.L. (2006) A nondiapausing variant of the flesh fly, *Sarcophaga bullata*, that shows arrhythmic adult eclosion and elevated expression of two circadian clock genes, *period* and timeless. *J Insect Physiol* **52**: 1213–1218.
Hardie, J. and Vaz Nunes, M. (2001) Aphid photoperiodic clocks. *J Insect Physiol* **47**: 821–832.
Hardin, P.E. (2005) The circadian timekeeping system of *Drosophila*. *Curr Biol* **15**: R714–R722.
Hillman, W.S. (1973) Non-Circadian Photoperiodic Timing in the Aphid Megoura. *Nature* **242**: 128–129.
Hirayama, J. and Sassone-Corsi, P. (2005) Structural and functional features of transcription factors controlling the circadian clock. *Curr Opin Genet Dev* **15**: 548–556.
Hirayama, J., Nakamura, H., Ishikawa, T., Kobayashi, Y. and Todo, T. (2003) Functional and structural analyses of cryptochrome. Vertebrate CRY regions responsible for interaction with the CLOCK:BMAL1 heterodimer and its nuclear localization. *J Biol Chem* **278**: 35620–35628.
Hodgson, C.J. and Lane, I.R. (1981) Some effects of photoperiod on larviposition and fresh weight-gain in *Myzus persicae*. *Physiol Entomol* **6**: 21–25.
Ikeno, T., Numata, H. and Goto, S.G. (2008) Molecular characterization of the circadian clock genes in the bean bug, *Riptortus pedestris*, and their expression patterns under long- and short-day conditions. *Gene* **419**: 56–61.
Ivanchenko, M., Stanewsky, R. and Giebultowicz, J.M. (2001) Circadian photoreception in *Drosophila*: functions of cryptochrome in peripheral and central clocks. *J Biol Rhythms* **16**: 205–215.
Koh, K., Zheng, X. and Sehgal, A. (2006) JETLAG resets the *Drosophila* circadian clock by promoting light-induced degradation of TIMELESS. *Science* **312**: 1809–1812.
Kotwica, J., Bebas, P., Gvakharia, B.O. & Giebultowicz, J.M. (2009) RNA interference of the period gene affects the rhythm of sperm release in moths. *J Biol Rhythms* **24**: 25–34.
Kyriacou, C.P. and Hall, J.C. (1980) Circadian rhythm mutations in *Drosophila melanogaster* affect short-term fluctuations in the male's courtship song. *Proc Natl Acad Sci USA* **77**: 6729–6733.
Kyriacou, C.P. and Hall, J.C. (1986) Interspecific genetic control of courtship song production and reception in *Drosophila*. *Science* **232**: 494–497.
Kyriacou, C.P., Peixoto, A.A., Sandrelli, F., Costa, R. and Tauber, E. (2007) Clines in clock genes: fine-tuning circadian rhythms to the environment. *Trends in Genet* **24**: 124–132.
Lees, A.D. (1973) Photoperiodic time measurement in the aphid *Megoura viciae*. *J Insect Physiol* **19**: 2279–2316.
Lin, C. and Todo, T. (2005) The cryptochromes. *Genome Biol* **6**: 220.
Lincoln, G.A., Andersson, H. and Loudon, A. (2003) Clock genes in calendar cells as the basis of annual timekeeping in mammals – a unifying hypothesis. *J Endocrinol* **179**: 1–13.
Looby, P. and Loudon, A.S. (2005) Gene duplication and complex circadian clocks in mammals. *Trends Genet* **21**: 46–53.
Markovitch, S. (1924) The migration of the Aphididae and the appearance of sexual forms as affected by the relative length of daily exposure. *J Agric Res* **27**: 513–522.
Mathias, D., Jacky, L., Bradshaw, W.E. and Holzapfel, C.M. (2007) Quantitative Trait Loci Associated with Photoperiodic Response and Stage of Diapause in the Pitcher-Plant Mosquito, *Wyeomyia smithii*. *Genet* **176**: 391–402.
Nakabachi, A., Shigenobu, S., Sakazume, N., Shiraki, T., Hayashizaki, Y., Carninci, P. *et al.* (2005) Transcriptome analysis of the aphid bacteriocyte, the symbiotic host cell that harbors an endocellular mutualistic bacterium, Buchnera. *Proc Natl Acad Sci USA* **102**: 5477–5482.
Narayandas, G.K. and Alyokhin, A.V. (2006) Diurnal patterns in host finding by potato aphids, *Macrosiphum euphorbiae* (Homoptera: Aphididae). *J Insect Physiol* **19**: 347–356.
Nelson, W., Tong, Y.L., Lee, J.K. and Halberg, F. (1979) Methods for cosinor-rhythmometry. *Chronobiologia* **6**: 305–323.
Pavelka, J., Shimada, K. and Kostal, V. (2003) TIMELESS: a link between fly's circadian and photoperiodic clocks? *Eur J Entomol* **100**: 255–265.
Peschel, N., Veleri, S. and Stanewsky, R. (2006) Veela defines a molecular link between Cryptochrome and Timeless in the

light-input pathway to *Drosophila*'s circadian clock. *Proc Natl Acad Sci USA* **103**: 17313–17318.

Pfaffl, M.W. (2001) A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res* **29**: 2002–2007.

Piccin, A., Couchman, M., Clayton, J.D., Chalmers, D., Costa, R. and Kyriacou, C.P. (2000) The clock gene *period* of the housefly, *Musca domestica*, rescues behavioural rhythmicity in *Drosophila melanogaster*: evidence for intermolecular coevolution? *Genet* **154**: 747–758.

Reddy, K.L., Wohlwill, A., Dzitoeva, S., Lin, M.H., Holbrook, S. and Storti, R.V. (2000) The *Drosophila* PAR domain protein 1 (Pdp1) gene encodes multiple differentially expressed mRNAs and proteins through the use of multiple enhancers and promoters. *Dev Biol* **224**: 401–414.

Robinson-Rechavi, M. and Huchon, D. (2000) RRTree: relative-rate tests between groups of sequences on a phylogenetic tree. *Bioinformatics* **16**: 296–297.

Rubin, E.B., Shemesh, Y., Cohen, M., Elgavish, S., Robertson, H.M. and 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 Res* **16**: 1352–1365.

Saez, L. and Young, M.W. (1996) Regulation of nuclear entry of the *Drosophila* clock proteins period and timeless. *Neuron* **17**: 911–920.

Sakai, T. and Ishida, N. (2001) Circadian rhythms of female mating activity governed by clock genes in *Drosophila*. *Proc Natl Acad Sci USA* **98**: 9221–9225.

Sakakida, Y., Miyamoto, Y., Nagoshi, E., Akashi, M., Nakamura, T.J., Mamine, T. *et al.* (2005) Importin alpha/beta mediates nuclear transport of a mammalian circadian clock component, mCRY2, together with mPER2, through a bipartite nuclear localization signal. *J Biol Chem* **280**: 13272–13278.

Sandrelli, F., Tauber, E., Pegoraro, M., Mazzotta, G., Cisotto, P., Landskron, J. *et al.* (2007) A molecular basis for natural selection at the *timeless* locus in *Drosophila melanogaster*. *Science* **316**: 1898–1900.

Sauman, I. and Reppert, S.M. (1996) Circadian clock neurons in the silkmoth *Antheraea pernyi*: novel mechanisms of Period protein regulation. *Neuron* **17**: 889–900.

Saunders, D.S., Lewis, R.D. and Warman, G.R. (2004) Photoperiodic induction of diapause: opening the black box. *Physiol Entomol* **29**: 1–15.

Schmidt, H.A., Strimmer, K., Vingron, M. and von Haeseler, A. (2002) TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* **18**: 502–504.

Schultz, J., Milpetz, F., Bork, P. and Ponting, C.P. (1998) SMART, a simple modular architecture research tool: identification of signaling domains. *Proc Natl Acad Sci USA* **95**: 5857–5864.

Shimada, K. (2005) Photoperiod-sensitive developmental delay in facet mutants of the drosophilid fly, *Chymomyza costata* and the genetic interaction with timeless. *J Insect Physiol* **51**: 649–653.

Staden, R., Beal, K. and Bonfield, J.K. (1998) The Staden package. In *Computer Methods in Molecular Biology* (Misener, S. and Krawetz, S., eds), pp. 115–130. Humana Press, Totowa, NJ.

Stanewsky, R. (2003) Genetic analysis of the circadian system in *Drosophila melanogaster* and mammals. *J Neurobiol* **54**: 111–147.

Stehlik, J., Zavodska, R., Shimada, K., Sauman, I. and Kostal, V. (2008) Photoperiodic induction of diapause requires regulated transcription of *timeless* in the larval brain of *Chymomyza costata*. *J Biol Rhythms* **23**: 129–139.

Stoleru, D., Nawathean, P., Fernandez, M.P., Menet, J.S., Ceriani, M.F. and Rosbash, M. (2007) The *Drosophila* circadian network is a seasonal timer. *Cell* **129**: 207–219.

Takahata, S., Ozaki, T., Mimura, J., Kikuchi, Y., Sogawa, K. and Fujii-Kuriyama, Y. (2000) Transactivation mechanisms of mouse clock transcription factors, mClock and mArnt3. *Genes Cells* **5**: 739–747.

Tamura, K., Dudley, J., Nei, M. and Kumar, S. (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* **24**: 1596–1599.

Tauber, E., Roe, H., Costa, R., Hennesy, J.M. and Kyriacou, C.P. (2003) Temporal mating isolation driven by a behavioural gene in *Drosophila*. *Curr Biol* **13**: 140–145.

Tauber, E., Zordan, M., Sandrelli, F., Pegoraro, M., Osterwalder, N., Breda, C. *et al.* (2007) Natural selection favors a newly derived *timeless* allele in *Drosophila melanogaster*. *Science* **316**: 1895–1898.

Thieme, T. and Dixon, A.F.G. (1996) Mate recognition in the *Aphis fabae* complex: daily rhythm of release and specificity of sex pheromones. *Entomol Exp Appl* **79**: 85–89.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. and Higgins, D.G. (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **24**: 4876–4882.

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

Vaz Nunes, M. and Hardie, J. (1993) Circadian rhythmicity is involved in photoperiodic time measurement in the aphid *Megoura viciae*. *Cell Mol Life Sci* **49**: 711–713.

Veerman, A. (2001) Photoperiodic time measurement in insects and mites: a critical evaluation of the oscillator-clock hypothesis. *J Insect Physiol* **47**: 1097–1109.

Williams, J.A. and Sehgal, A. (2001) Molecular components of the circadian system in *Drosophila*. *Annu Rev Physiol* **63**: 729–755.

Yuan, Q., Metterville, D., Briscoe, A.D. and Reppert, S.M. (2007) Insect cryptochromes: gene duplication and loss define diverse ways to construct insect circadian clocks. *Mol Biol Evol* **24**: 948–955.

Zhu, H., Yuan, Q., Briscoe, A.D., Froy, O., Casselman, A. and Reppert, S.M. (2005) The two CRYs of the butterfly. *Curr Biol* **15**: R953–R954.

Zhu, H., Sauman, I., Yuan, Q., Casselman, A., Emery-Le, M., Emery, P. *et al.* (2008) Cryptochromes define a novel circadian clock mechanism in monarch butterflies that may underlie sun compass navigation. *PLoS Biol* **6**: e4.

Supporting Information

Additional Supporting Information may be found in the online version of this article under the DOI reference: DOI 10.1111/j.1365-2583.2009.00931.x

Figure S1. Phylogenetic relationships among insect paralogous genes *Clock, Cycle* and *Arntl/Tango*. A neighbor-joining topology resulting from the analysis of Poisson-corrected distances is shown with bootstrap values indicated when higher than 50%. For each gene, the *Daphnia* sequence was also included. The tree obtained shows the true orthology for each of sequences found in the *Acyrthosiphon pisum* genome. The anomalous positions of the pea aphid sequences, basal to *Daphnia* and the rest of insects, specially in the *Cycle* clade, further reflects the accelerated rates for these genes in *A. pisum* shown in Table 1.

Figure S2. Phylogenetic relationships among animal cryptochromes and the 6-4 Photolyase, including homolog sequences found in *Acyrthosiphon pisum*. Both Insect CRY1 (*Drosophila*-type) and Insect CRY2 (mammal-type) are present in the *A. pisum* genome, with two copies found for the latter (named AcpCRY2-1 and AcpCRY2-2). The phylogenetic inference was done using the neighbour-joining algorithm on Poisson-corrected distances. Bootstrap values are shown for nodes when higher than 50%. The 6-4 Photolyase of *Arabidopsis thaliana* was used to root the tree.

Table S1. Primers used for PCR amplification, sequencing and Real-time quantitative PCR of the genes characterized in the present study.

Table S2. Accession numbers for sequences included in the phylogenetic analysis.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.