Tropical Plant Biol. (2008) 1:236–245
DOI 10.1007/s12042-008-9020-3

Conserved Daily Transcriptional Programs in *Carica papaya*

Anna Zdepski · Wenqin Wang · Henry D. Priest · Faraz Ali · Maqsudul Alam · Todd C. Mockler · Todd P. Michael

Received: 22 September 2008 / Accepted: 12 November 2008 / Published online: 16 December 2008

Abstract Most organisms have internal circadian clocks that mediate responses to daily environmental changes in order to synchronize biological functions to the correct times of the day. Previous studies have focused on plants found in temperate and sub-tropical climates, and little is known about the circadian transcriptional networks of plants that typically grow under conditions with relatively constant day lengths and temperatures over the year. In this study we conducted a genomic and computational analysis of the circadian biology of *Carica papaya*, a tropical tree. We found that predicted papaya circadian clock genes cycle with the same phase as *Arabidopsis* genes. The patterns of time-of-day overrepresentation of circadian-associated promoter elements were nearly identical across papaya, *Arabidopsis*, rice, and poplar. Evolution of promoter structure predicts the observed morning- and evening-specific expression profiles of the papaya *PRR5* paralogs. The strong conservation of previously identified circadian transcriptional networks in papaya, despite its tropical habitat and distinct life-style, suggest that circadian timing has played a major role in the evolution of plant genomes, consistent with the selective pressure of anticipating daily environmental changes. Further studies could exploit this conservation to elucidate general design principles that will facilitate engineering plant growth pathways for specific environments.

Keywords *Carica papaya* · Circadian clock · *Cis*-acting element · Diurnal

Abbreviations

Cp	*Carica papaya*
At	*Arabidopsis thaliana*
Os	*Oryza sativa*
Pt	*Populus trichocarpa*
bp	*basepair*
PRR	PSEUDO-RESPONSE REGULATOR

Introduction

Almost all organisms across kingdoms have internal biological clocks mediating annual, seasonal and daily changes in the environment [21]. The circadian clock controls daily biological rhythms and confers fitness to an organism by synchronizing internal biology with that of the rhythmic external environment [6, 27]. The circadian clock is an endogenous self-sustaining timing mechanism with a period of approximately 24 h that can be entrained to the exact timing of daily environmental cycles over a range of physiologically relevant temperatures. In addition, in plants

Anna Zdepski, Wenqin Wang, Henry D. Priest have contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s12042-008-9020-3) contains supplementary material, which is available to authorized users.

A. Zdepski · W. Wang · F. Ali · T. P. Michael (✉)
Department of Plant Biology and Pathology,
The Waksman Institute of Microbiology, Rutgers,
The State University of New Jersey,
Piscataway, NJ, 08854, USA
e-mail: tmichael@waksman.rutgers.edu

H. D. Priest · T. C. Mockler
Department of Botany and Plant Pathology and Center
for Genome Research and Biocomputing,
Oregon State University,
Corvallis, Oregon 97331, USA

M. Alam
Department of Microbiology, University of Hawaii,
Honolulu, Hawaii 96822, USA

there are at least two circadian clocks that can be distinguished based on their ability to synchronize with light or temperature [25].

The circadian clocks of bacteria, fungi, plants, and animals are thought to have evolved independently, but all are comprised of negative feedback loops of transcription and regulated protein turnover [21]. The current model of the plant circadian clock has been worked out in *Arabidopsis thaliana* and consists of three interlocking feedback loops [19, 40]. Three protein families define the plant circadian clock with unique combinations of domains conserved across multiple species: single MYB transcription factors (sMYB); pseudo-response regulators with a CONSTANS domain (PRR/CCT); and PAS/FBOX/KELCH (PFK). In addition there are multiple proteins that play a role in the circadian clock, light signaling and flowering time: GIGANTEA (GI); EARLY FLOWERING 3 (ELF3); ELF4; TIME FOR COFFEE (TIC); TEJ; and casein kinase beta subunit 3 (CKB3) [21].

The two homologous morning-expressed single Myb transcription factors *CIRCADIAN CLOCK ASSOCIATED 1* (*CCA1*) [39] and *LATE ELONGATED HYPOCOTYL* (*LHY*) [35] repress the expression of the pseudo-response regulator *TIMING OF CAB2 EXPRESSION 1* (*TOC1*) [37] and *LUX ARRYTHMO* (*LUX*) by binding to evening elements (EE) in their promoters [13]. *TOC1* and *LUX* levels increase toward the end of the day and directly or indirectly up-regulate expression of *CCA1* and *LHY* [1, 13]. Again, through binding of the EE, CCA1 and LHY activate the expression of their repressors, pseudo-response regulators *PRR7* and *PRR9* [9]. The third interlocked loop, *CCA1*, *LHY*, and *TOC1* act to repress the likely activator of *TOC1*, *GI* [19].

Based on microarrays, 89% of the *Arabidopsis* transcriptome is expressed at different levels over the day depending on the environmental conditions and for any specific condition 15–30% of transcripts cycle under circadian conditions while 30–50% cycle under diurnal conditions [2, 5, 7, 22]. These numbers are consistent with an estimate that 35% of the transcriptome is circadian regulated based on enhancer trapping [24, 30]. Processes such as growth are controlled by time-of-day coordination of phytohormone expression pathways by the circadian clock and light signaling [26]. Furthermore, three time-of-day specific transcriptional modules were identified that are conserved across *Arabidopsis*, poplar and rice [22]. This latter finding suggests that daily environmental cycles have contributed significantly to shaping the fabric of the plant genome.

One of the driving questions of the current study was whether or not circadian timing would be conserved between *Arabidopsis* and papaya because as a tropical plant papaya primarily grows at latitudes with less seasonal variation in day length and temperature than *Arabidopsis*. Papaya has about half the circadian clock and light signaling genes as *Arabidopsis*, yet an expansion of the *COP1* gene family that mediates degradation of light signaling proteins. This led to the hypothesis that papaya may spend less energy measuring time and more energy degrading proteins in direct response to changes in light [28]. This later hypothesis would be consistent with the idea that synchronous flowering near the equator is governed by the perception of variation in sunrise and sunset [3]. In this study we address the question of whether the circadian clock is conserved in papaya, a predominately tropical plant. We find that the transcriptional networks and expression are conserved in papaya. The results presented here suggest that circadian timing has played a major role in the evolution of plant genomes.

Results

Carica Papaya Circadian Clock and Light Signaling Orthologs

The draft *Carica papaya* genome sequence provided an opportunity to investigate a tropical circadian clock at the molecular level [28]. As a first approximation of the gene content of papaya we used a protein mutual best-blast match strategy to identify putative orthologs with *Arabidopsis* (Material and Methods). We focused on this comparison as *Arabidopsis* circadian clock research provides the most extensive information at the molecular and genetic level. Putative papaya-rice, papaya-poplar and papaya-sorghum orthologs were identified and included in a searchable database called ORTHOMAP (http://orthomap.cgrb.oregonstate.edu/).

Many circadian clock and light signaling gene families are smaller in papaya compared to *Arabidopsis*, rice and poplar, which is consistent with the lack of genome duplication in papaya [28]. For instance, there is only one homolog in papaya of the PAS-PAC/FBOX/KELCH (PFK) family gene ZTL compared to three in *Arabidopsis* (ZTL, FKF1 and LKP2; Table 1). In addition, there is only one homolog of the single MYB transcription factor LHY compared to two paralogs in *Arabidopsis* (LHY and CCA1; Table 1).

In contrast, papaya has the five pseudo-response regulators (PRRs) just like *Arabidopsis* [28]. Since most circadian genes are reduced in papaya compared to *Arabidopsis* and the PRRs were not, we took a closer look at the PRR gene family in papaya. Similar to the situation in rice, which also has five PRR proteins [33], the papaya PRRs can be separated into three groups: *PRR1/TOC1*, *PRR5/9* and *PRR7/3* (Fig. 1). As in *Arabidopsis* and rice, there is only one *PRR1/TOC1* gene in papaya that we designated *CpPRR1*. To date, fully sequenced plant

Table 1 *Carica papaya* light and circadian orthologs

Name	Domain (SMART)	Papaya gene ID	AtBestBLAST	At gene	Function
ZTL	PAS/FBOX (no Kelch)	supercontig_95.43	At5g57360	ZTL	clock, light signaling
PHOT1	PAS/PAC/PAS/PAC/STYKc	supercontig_139.19	At3g45780	PHOT1	light signaling
LHY	sMYB-A	supercontig_57.78	At1g01060	LHY	clock, light signaling
RVE1	sMYB-A	supercontig_178.22	At5g17300	RVE1	NA
ERP1	sMYB-A	supercontig_7.134	At1g18330	EPR1, RVE7	light signaling
RVE6	sMYB-A	supercontig_114.57	At3g09600	RVE6	NA
LUX	sMYB-B	supercontig_81.106	At3g46640	LUX	clock
LUX4	sMYB-B	supercontig_92.70	At3g10760	LUX4	NA
PRR5A	PRR/CCT	supercontig_3.152	At5g24470	PRR5	clock, light, flowering
PRR5B	PRR/CCT	supercontig_193.20	At5g24470	PRR5	clock, light, flowering
PRR7A	PRR/CCT	supercontig_1.291	At5g02810	PRR7	clock, light, flowering
PRR7B	PRR/CCT	supercontig_139.32	At5g02810	PRR7	clock, light, flowering
TOC1	PRR/CCT	supercontig_13.294	At5g61380	TOC1, PRR1	clock, light, flowering
TEJ	PARG	supercontig_9.247	At2g31840	Unknown;PARG-like	clock
CKB3	CASEIN KINASE II	supercontig_98.66	At4g17640	CKB2	clock
GI	UKNOWN	supercontig_26.82	At1g22770	Gl	clock, light, flowering
ELF3	UKNOWN	supercontig_78.11	At2g25930	ELF3	clock, light gating
ELF4A	UKNOWN	NA	At2g40080	ELF4	clock, light gating
ELF4B	UKNOWN	NA	At2g40080	ELF4	clock, light gating
ELF4-L3	UKNOWN	NA	At2g06255	ELF4-L3	NA
ELF4-L4	UKNOWN	supercontig_25.111	At1g17455	ELF4-L4	NA
SRR1	UKNOWN	supercontig_30.42	At5g59560	SRR1	clock, light
TIC	UKNOWN	supercontig_58.130	At3g22380	TIC	Clock
TKL	UKNOWN	NA	At3g63180	TKL	NA

genomes contain only one copy of the *PRR1/TOC1* gene if at all (TP. Michael, unpublished observations). In papaya, neither *PRR9* nor *PRR3* had mutual best-blast orthologs based on our criteria (Material and Methods). In contrast, both *PRR5* and *PRR7* had mutual best-blast matches in addition to closely related homologues (one way blast), which also clustered with *PRR9* and *PRR3* based on multiple alignment (Fig. 1). In rice, due to this close relationship and the inability to separate the two *PRR5/PRR9* and *PRR7/PRR3* paralogs, they were named *OsPRR5/9*, *OsPRR9/5, OsPRR7/3 and OsPRR3/7* [32]. Based on our mutual best-blast criteria we chose to name the papaya *PRR* genes *CpPRR5A*, *CpPRR5B*, *CpPRR7A* and *CpPRR7B* where A and B represent paralogous proteins. Regardless of the our naming strategy, our results are consistent with a trend in the expansion of the *PRR* gene family across species where three clades emerged and specific members in at least two clades, *PRR5*/9 and *PRR7*/3, expanded.

Conserved Time-of-day *Cis*-Acting Elements

In *Arabidopsis* there are three *cis*-acting modules controlling time of day expression: the morning module, morning element (ME, CCACAC)/Gbox (CACGTG); the evening module, evening element (EE, AAATATCT)/GATA (GATA); and the midnight module, telobox (TBX, AAACCCT)/starch synthesis box (SBX, AAGCCC)/ protein box (PBX, ATGCCC) [22]. These three modules are also conserved across divergent species such as poplar and rice, suggesting that time-of-day signaling has specifically shaped the evolution of transcriptional networks in higher plants, and possibly photosynthetic organisms in general [22].

To address whether these time-of-day *cis*-acting modules are conserved in papaya, we assigned the phase of expression from the *Arabidopsis* orthologs to papaya and searched for time-of-day specific overrepresented elements in the promoters of the papaya orthologs [22]. We utilized the phase of *Arabidopsis* genes from eight diurnal and circadian conditions. Every 3–8 bp word from 500 bp of papaya promoters was queried for overrepresentation in each of the phase-specific gene lists using the ELEMENT promoter-searching tool [30]. For each word, plotting the Z-score at each phase over the day generated a Z-score profile. Only words with Z-scores that were significant at more than two consecutive phases over the day were retained in the analysis (Material and Methods). Across eight diurnal and circadian conditions there were between 250 and 578 significant words (Fig. 2a). We clustered significant 3–8mer words based on the time-of-day (hrs from lights-on/subjective dawn) that the Z-score most significant (highest peak), i.e. the words were grouped by the time-of-day that they were most overrepresented and presumably active. We found a similar number of words at

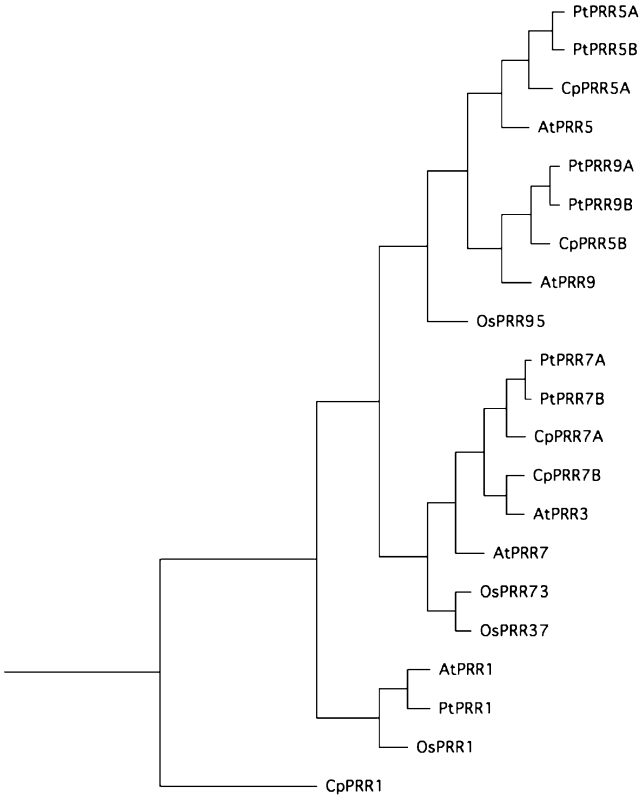

Fig. 1 Three PRR gene branches in papaya PRR family: PRR1, PRR5 (A and B) and PRR7 (A and B). *Arabidopsis thaliana* (At), *Populus trichocarpa* (Pt), *Orzyza sativa* (Os) and *Carica papaya* (Cp) PRR proteins were aligned with clustalX and the tree was constructed with TreeView

each time of the day (phase), except in two conditions where we found more words later in the night (Fig. 2b).

We took a closer look at the words that were overrepresented around midnight to early morning. We found that many (30–55%) of the words could be summarized into two elements, the ME and TBX (Fig. 3), which we have previously identified in *Arabidopsis*, poplar and rice [22]. The remaining words similar to words that we have found previously in *Arabidopsis* (discussed below), while other words are specific to papaya and may represent novel papaya specific elements. However, we did note words that were "AT" rich that seem to be specific to papaya. These words could represent a new class of time-of-day specific elements, or could be an artifact of our *in silico* analysis. More experimentation in papaya will be required to resolve these possibilities.

The words that make up the ME were overrepresented late at night and at the beginning of the day: phases 22, 23, 24/0, 1 and 2. In contrast, the words that make up the TBX were overrepresented around midnight, phases 17, 18, 19 and 20. When we grouped overrepresented words over the day by phase, and then aligned the words based on sequence similarity, we were able to summarize them into consensus words or elements. We think of the summary element as related to a transcription factor binding site, or *cis*-element. We then plotted out the summarized elements, grouped them by time-of-day of peak Z-score significance and compared their Z-score profiles to those obtained for *Arabidopsis*, rice and poplar (Fig. 4, Material and Methods). Consistent with our previous results [22], the pattern of overrepresentation (Z-score profile) for the TBX, Gbox, ME, and EE shared consistent time-of-day Z-score peaks across these distantly related species. In addition, we found that the conservation of the Z-score profile was highly specific across conditions. For instance, in *Arabidopsis* we found that the phase of overrepresentation for the TBX is dependent on condition [22]. Similar to *Arabidopsis*, in papaya we found that under conditions without temperature

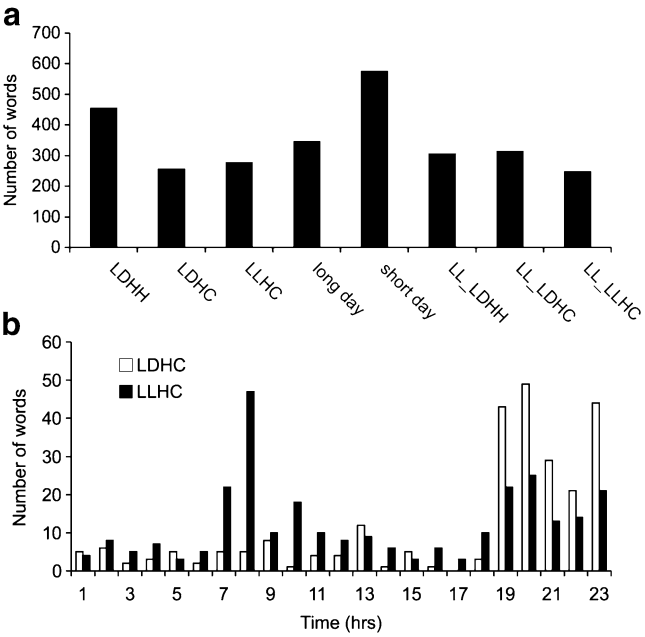

Fig. 2 Time-of-day words (3–8mers) identified in papaya. **a** Number of time-of-day specific words identified per condition in papaya promoters. Best-blast orthologs were identified between *Arabidopsis*-papaya, the phase of *Arabidopsis* genes was assigned to the papaya ortholog, Z-scores were calculated for every 3–8mer from 500 bp of the papaya promoter by looking for enrichment over observed in a similar sample size from the genome, Z-scores were plotted over the day and only words with two consecutive Z-scores above the threshold (~3, $P<0.05$) were retained. Between 200 and 600 words were retained across the eight conditions tested. Conditions are described [22]; LDHH: 12 h Light (L)-12 h dark (D) and continuous temperature (HH); LDHC: 12 h Light (L)-12 h dark (D) and 12 h hot (H)- 12 h cold (C); LLHC: continuous light (LL) and 12 h H-12 h C; long day: 16 h light-8 h dark and continuous temperature; short day: 8 h light-16 h dark and continuous temperature; LL_LDHH, LL_LDHC and LL_LLHC are sampled under continuous light (LL) and grown under the specified condition. **b** Words arranged by their time-of-day of overrepresentation (phase). Words were arranged by the time-of-day for two or more consecutive Z-scores that were greater than the Z-score threshold (~3, $P<0.05$). Two conditions LDHC (*white*) and LLHC (*black*) are plotted as representative of the eight conditions tested

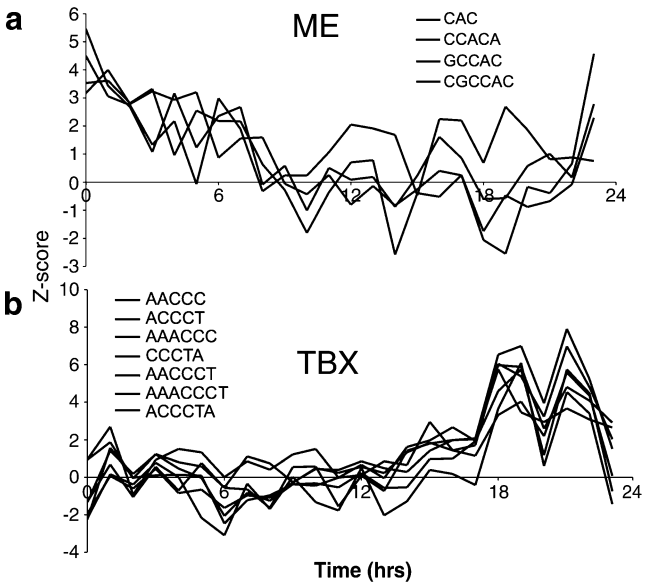

Fig. 3 The morning element (ME) and telobox (TBX) are predicted to be active at dawn and midnight, respectively. Words that were overrepresented under the condition LDHC with the same phase were grouped by sequence similarity. **a** The morning element (ME: CCACAC) was overrepresented at dawn. Multiple words CAC, CCAC, GCCAC, and CGCCAC were summarized as having dawn-specific overrepresentation. **b** The telobox (TBX: AAACCCT) was overrepresented around midnight. Multiple words AACCC, ACCCT, AAACCC, CCCTA, AACCCT, AAACCCT and ACCCTA were summarized as having midnight overrepresentation. Z-score threshold ~3, $P<0.05$

cycles, the TBX is overrepresented before dusk, while under any condition that includes a temperature cycle the TBX is overrepresented before dawn (Fig. 5). This finding is consistent with circadian transcriptional networks being highly conserved across these distantly related species.

Predicted Circadian Clock Genes Cycle Under Intermediate Day Conditions in *C. Papaya*

One assumption underpinning comparative genome analysis is that orthologous genes behave in a similar way across species. To date, the circadian clock of *Arabidopsis* is the best described in higher plants. Multiple groups have studied the expression of circadian clock orthologs to determine if the timing of expression is conserved across species [4, 8, 14, 18, 29, 32-34]. In almost all cases, the phase of expression is conserved across species (Table 2).

To test this in papaya, we performed two independent 48-hour time courses in mature papaya trees. We collected young emerging leaves and measured expression by quantitative real-time PCR (qPCR). We found that all of the predicted circadian clock orthologs cycle with the same phase in papaya as in *Arabidopsis* under both diurnal and circadian conditions (Table 2; Fig. 6, Fig. S1). *CpLHY* peaks in the morning while *CpTOC1* peaks in the evening under both diurnal and circadian conditions (Fig. 6, Fig. S1). These results are consistent with our findings that the transcriptional networks are conserved between papaya and *Arabidopsis*, and also consistent with the idea that time-of-day networks may be conserved across higher plants. In addition, we tested our Arabidopsis-papaya ortholog phase predictions on a several randomly selected genes and found that the phase of expression was the same in papaya as in *Arabidopsis* (Fig. S2). Together with the *in silico* promoter analysis, these findings support global conservation of time-of-day expression in papaya.

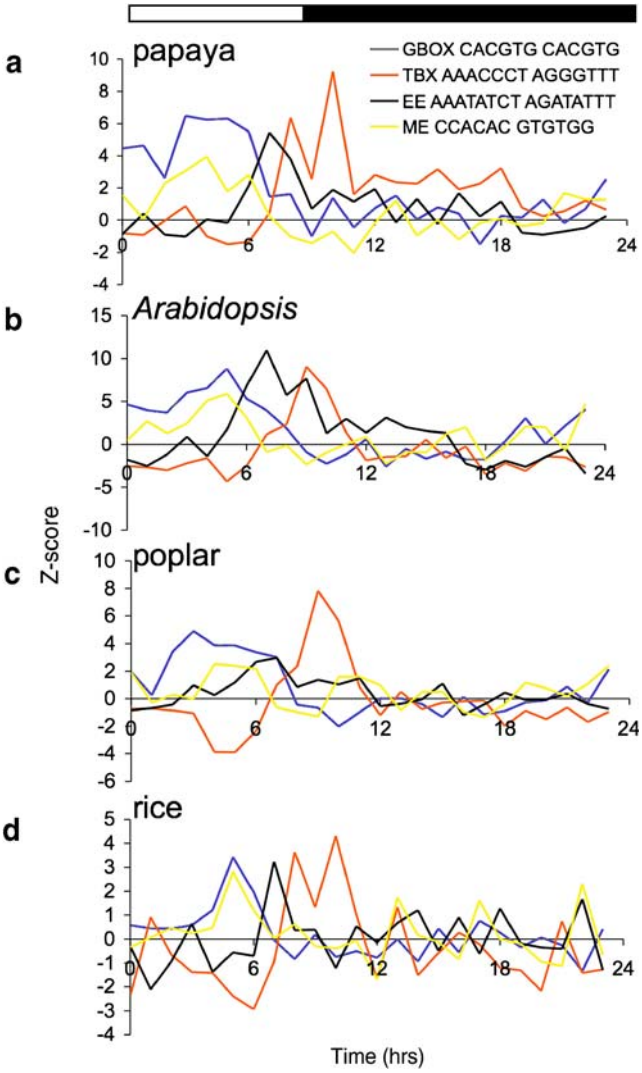

Fig. 4 Time-of-day overrepresentation circadian *cis*-elements conserved in papaya, *Arabidopsis*, poplar and rice. Z-score profiles were summarized into elements for words sharing both sequence similarity and time-of-day overrepresentation. Z-score profiles for the GBOX (*blue*, CACGTG), TBX (*orange*, AAACCCT), EE (*black*, AAATATCT) and ME (*yellow*, CCACAC) were conserved across papaya **a**, *Arabidopsis* **b**, poplar **c** and rice **d**. Z-score threshold ~3, $P<0.05$. The reverse complement of each element is presented in the legend

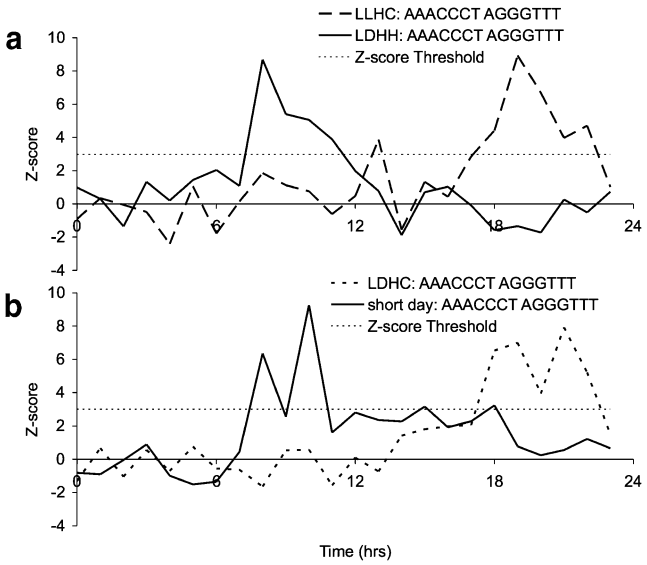

Fig. 5 The papaya TBX displays a condition dependent Z-score profile shift. The papaya TBX (AAACCCT) Z-score profile displays a distinct time-of-day overrepresentation depending on the *Arabidopsis* condition used to assign phase to papaya. **a** LLHC (*dotted line*) and LDHH (*solid line*); **b** LDHC (*dotted line*) and short day (*solid line*); Z-score threshold (thin dotted line, ~3, $P<0.05$). Conditions explained in Fig. 2 legend

However, in rice, the phasing of some of the *PRR* family members is distinct from their *Arabidopsis* orthologs. Whereas *AtPRR9* peaks at dawn and *AtPRR5* peaks in the late afternoon, *OsPRR95* and *OsPRR59* peak at the same time in the evening. We found that *CpPRR5A* and *CpPRR5B*, similar to *AtPRR5* and *AtPRR9* respectively, displayed distinct time of day expression (Fig. 7a). *CpPRR5A* peaked in the late afternoon/dusk, similar to *AtPRR5*, while *CpPRR5B* peaks in the morning, similar to *AtPRR9*. The expression patterns of the papaya *PRR5* genes is consistent with the multiple alignment in Fig. 1, which suggests that *CpPRR5B* or more closely related to *AtPRR9.*

We were very interested in the phasing of the *CpPRR5* paralogs, so we looked more closely at the *CpPRR5* promoter regions. We compared 500 bp from *CpPRR5A* and *CpPRR5B* and found a concentration of circadian response elements (EE/GATA, ME/Gbox) and the TGA element (TGACGTGG), which is found in multiple copies in both *Arabidopsis* [24] and poplar *PRR* promoters (T.P. Michael, unpublished observation). Both promoter regions were highly similar for circadian response element position and number, with two exceptions (Fig. 7b). We found that there was an expansion of an EE cluster in the *CpPRR5A* promoter and an ME expansion in the *CpPRR5B* promoter. Based on our promoter analysis here and published [22, 24], and empirical *in vivo* promoter studies [12, 23], we would predict that an increased number of EE or ME would confer evening-specific or morning-specific expression respectively. Therefore, the difference in timing of expression between *CpPRR5A* and *CpPRR5B* is consistent with their promoter composition. Moreover, we found it striking that the expansions of the EE and ME occurred without much disruption to the linear arrangement of elements in the papaya *PRR5* promoters. This suggested to us that the papaya *PRR* promoters may have diverged very recently and that there was some selective pressure to maintain the expression difference between these two genes. We hy-

Table 2 Time-of-day (phase) expression of circadian clock orthologs across species

	Arabidopsis thaliana	*Carica papaya*	*Oryza sativa* (rice)	*Lemna paucicostata* (duckweed)	*Lemna gibba* (duckweed)	*Pisum sativum* (pea)	*Castanea sativa* (European chestnut)	*Solanum lycopersicum* (tomato)	*Mesembryanthemum crystallinum* (common ice plant)	*Glycine max* (soybean)	*Phaseolus vulgaris* (common bean)
CCA1	M	–	M	–	–	M	–	–	M	M	–
LHY	M	M	–	M	M	–	M	M	–	M	M
PRR1/TOC1	E	E	E	E	E	E	E	–	E	E	–
PRR3	E	D	D	D	D	–	–	–	–	–	–
PRR5	E	E	E	E	E	–	–	–	–	–	–
PRR7	D	E	D	–	–	–	–	E	–	–	–
PRR9	M	M	E	E	E	–	–	–	–	–	–
ELF3	N	E	–	E	E	–	–	–	E	–	–
ELF4	E	–	–	–	–	E	–	–	E	–	–
GI	E	E	E	E	E	E	–	–	E	–	–
FKF1	E	–	E	–	–	–	–	E	E	–	–
ZTL	NR	E	–	–	–	–	–	–	E	–	–
LUX	E	–	E	–	–	–	–	–	–	–	–

– not reported
M morning, *D* day, *E* evening, *N* night, *NR* not rhythmic

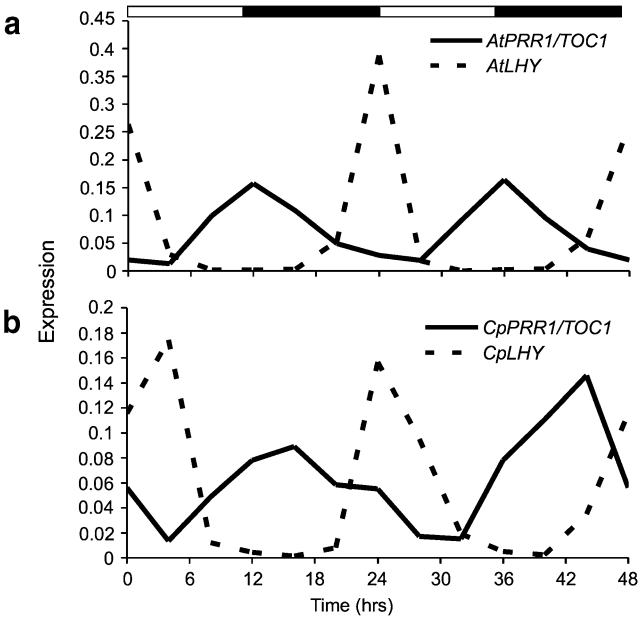

Fig. 6 Papaya circadian clock genes cycle with the same phase as *Arabidopsis* genes. Papaya circadian clock orthologs were used to design quantitative real-time PCR (qPCR) primers to detect relative transcript levels. New leaves from three-month-old trees were sampled under diurnal conditions every four hours over two days. **a** Diurnal expression of *AtTOC1/PRR1* (*solid black line*) and *AtLHY* (*dotted black line*). **b** Diurnal expression of *CpTOC1/PRR1* (*solid black line*) and *CpLHY* (*dotted black line*). *At, Arabidopsis thaliana* and *Cp, Carica papaya*. Data represent two independent biological replicates. Bars at top represent diurnal cycle (light, white box and dark, black box)

pothesize that in addition to the post-translational regulation reported for the members of the *PRR* family [10, 11, 15] time-of-day expression of the *PRR* family may represent another important layer of biological regulation, which is a substrate for selective pressure.

Discussion

In this study we utilized the draft papaya genome sequence to elucidate conservation of the circadian clock across distantly related plant species. We established a searchable database of papaya orthologs between *Arabidopsis*, poplar and rice and utilized this database to identify papaya circadian clock orthologs. We suggest that the *PRR* gene family in papaya reflects a recent gene family expansion compared with that in *Arabidopsis*. This was further verified with the functional diversification we found in the *CpPRR5* promoters where time-of-day specific elements have expanded to specify different phases of expression. We defined time-of-day *cis*-elements using a novel *in silico* promoter searching technique and demonstrated that time-of-day elements are conserved between papaya, *Arabidopsis*, poplar, and rice, consistent with selection acting directly on time-of-day activities. Finally, we demonstrate that circadian expression timing as well as sequence is conserved in circadian clock genes in papaya.

Almost universally across species the circadian clock orthologs of *CCA1/LHY* and *TOC1/PRR1* peak in the morning and evening respectively [4, 8, 14, 18, 29, 32–34]. Papaya is no exception and many genes cycle with similar phases as other species, suggesting that much of the basic circadian machinery in papaya is conserved. Differences between species do exist, both *CpZTL* and *McZTL* cycle with peak gene expression in the evening [4] (Table 2). In contrast, in *Arabidopsis* ZTL protein abundance cycles while its transcript does not [16, 36]. Considering the importance of ZTL in the circadian system [17], it will be interesting to see how *ZTL* expression impacts its functions in these other plant species.

The expression differences that we observed between the *PRR* family of genes in papaya, *Arabidopsis* and rice suggested that these expression differences contribute to functionality. In contrast to *PRR59* and *PRR95*, which have the same phase of expression in both duckweed and rice [29, 33], the gene expression of the paralogs in both papaya and *Arabidopsis* display distinct phases at midday and dusk. Similarly, the gene expression of the *PRR7/3* paralogs in papaya and *Arabidopsis* displays distinct phasing. This could represent a distinct functionality between monocots and dicots. In monocots, *PRR7* plays a prominent role in flowering time and has been identified in two different QTL studies with barley and rice [32, 38]. In contrast, QTL or induced mutants in *AtPRR7* result in modest changes in the

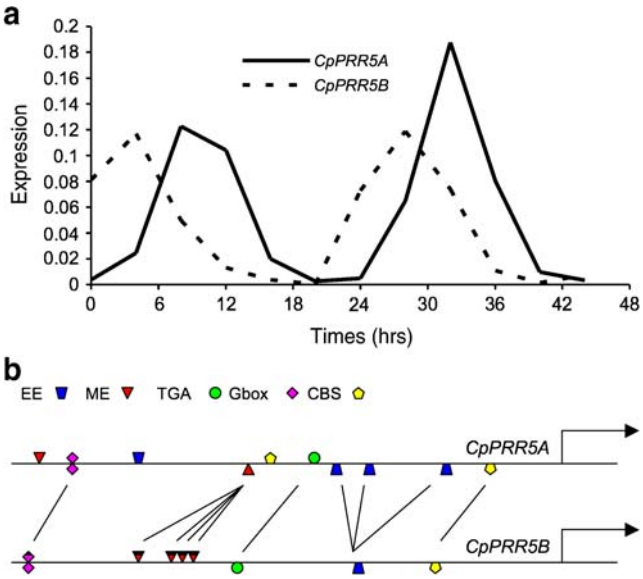

Fig. 7 Papaya *PRR5* promoter structure evolution predict morning and evening expression for *CpPRR5A* and *CpPRR5B* respectively. **a** Expression of *CpPRR5A* (*solid black line*) and *CpPRR5B* (*dotted black line*). **b** *CpPRR5A* and *CpPRR5B* promoter structure reveals similar spacing and expansion of known elements EE (*blue*), ME (*red square*), TGA (*green triangle*), Gbox (*pink*) and CBS (*yellow square*). ATG represented by arrow

circadian clock [9, 27]. The fact that the *CpPRR5A* and *CpPRR5B* promoters are similar in their linear arrangement of elements, and that increased numbers of specific *cis*-elements correlate with their distinct time-of-day expression, suggests that there must be some pressure to cause the expression of these genes to diverge from that found in the more basal monocots. This finding suggests that despite the importance of post-translational modification on the PRR proteins [10, 11, 15,], there must be evolutionary pressure on the temporal expression on these genes.

Consistent with idea that there is evolutionary pressure at the level of temporal expression, we have identified conserved *cis*-acting elements across papaya, *Arabidopsis*, rice and poplar using an *in silico* approach. These results extend our previous results between *Arabidopsis*, rice, and poplar [22], further confirming the power of this technique to identify conserved non-coding sequence between species and elucidate time-of-day transcriptional networks across plants. Recently we have verified these *cis*-elements empirically in both rice and poplar (T.C. Mockler and T.P. Michael, unpublished data). The fact that the *in silico* technique works between distantly related species with limited genomic colinearity suggests that there are groups of genes whose time-of-day co-expression is essential to plant fitness. Therefore, the genes that fall out of these co-expression clusters may represent novel activities in these species, providing a platform for identify diverging classes of genes between species. Time-of-day expression profiling facilitates the annotation of non-coding sequence and identification of novel functional gene clusters in newly sequenced plant species.

It was somewhat surprising that the time-of-day networks were conserved in papaya considering its tropical habitat and distinct life-style from *Arabidopsis*, rice and poplar. Yet it may be that the mechanism controlling synchronous flowering in tree at the equator is more of a circadian regulated system [3], and that the findings in temperate plants are broadly applicable. This conservation across distantly related plant species suggests that we could generate general principles that would apply to a host of plants for altering plant growth pathways for specific environments. For instance, one of these conserved time-of-day networks coordinates the growth promoting expression of phytohormones to coincide with the important environmental signal of the rising sun at dawn [26]. Recently, a forward genetic screen in the green algae *Chlamydomonas reinhardtii* revealed mutants in key plant circadian clock homologues, which cycle with similar phase of expression as *Arabidopsis* [20]. If time-of-day pathways are conserved across dicots, monocots, and single celled algae, this could provide new opportunities to engineer generalized strategies to control growth in an environment specific fashion from algae to higher plants.

Methods

Papaya Growth Conditions

Carica papaya transgenic variety ‘SunUp’ seeds were germinated and grown under intermediate days (12 h of light and 12 h dark, 12L12D) and continuous temperature (22°C) for three months to maturity. For the time courses, half of the trees were moved to a Continuous Light (LL) chamber at time 0 (T=0 h), lights on. The first leaf tissue collection began at T=24 h under continuous light and T=0 h under light/dark cycles. Tissue was collected every 4 h for two days under both circadian and diurnal conditions. Samples were immediately frozen in liquid nitrogen and stored at −80 C prior to RNA extraction.

Quantitative Real-time PCR

Quantitative realtime PCR (qPCR) was carried out as described [31]. Briefly, frozen papaya tissue was ground in 2 ml tubes with ball bearings. RNA was extracted using RNeasy (QIAGEN) with on column DNAase treatment, first strand cDNA was synthesized (Invitrogen, Carlsbad, CA) and used directly for qPCR assay on a myIQ (BioRad). Expression values were calculated as a function of CT values normalized to a standard dilution series over all samples assayed. Papaya primer sequences are described in Table S1.

Arabidopsis-Papaya Circadian Clock Orthologs

Arabidopsis-papaya orthologs were identified using a mutual best-blast hit strategy as described for *Arabidopsis*-rice and *Arabidopsis*-poplar [22]. In brief, papaya protein Y was blasted against all *Arabidopsis* proteins, which yields protein X as its best blast match, and then blasting protein X against all papaya proteins yields protein Y as protein X’s best-blast match. In this case, the two proteins are called best-blast matches (BBM) and referred to as putative orthologs. We further imposed a filter requiring all blast matches to be less than 1e-5, to reduce spurious poor (but still mutual best) blast matches. An *Arabidopsis*-papaya ortholog pair represents two proteins, which are mutual best blast matches. *Arabidopsis*-papaya, poplar-papaya, rice-papaya and sorghum-papaya orthologs can be searched using our online tool called ORTHOMAP (http://orthomap.cgrb.oregonstate.edu/).

Papaya *In Silico* Promoter Analysis

Papaya *in silico* promoter analysis was carried out as described for poplar and rice [22]. In brief, the phase of the best-blast *Arabidopsis* ortholog was assigned to its papaya

ortholog, and papaya gene lists for each phase of the day (0–23 hrs) were assembled. Each phase gene list contained hundreds of papaya genes and served as the input for promoter searching using ELEMENT (http://element.cgrb.oregonstate.edu) [30]. ELEMENT stores 500, 1000, 2000 and 3000 bp upstream of the predicted ATG from each papaya gene as putative promoter sequence; for this study we used 500 bp as the papaya promoter length. Using the papaya *in silico* phase gene lists, ELEMENT was used to assign an overrepresentation score, Z-score, for each 3–8mer in the papaya promoters; there are a total of 43,847 3–8mers. Every 3–8mer was assigned a Z-score for every phase of the day and was plotted as a function of time. We refer to the resulting graph as a "Z-score profile." The significance level of the Z-score profile was established as described [22]. Briefly, ELEMENT was used to assign a significance z-score to each word for each phase bin. The z-scores were then plotted for each phase bin over the day creating a 'z-score profile' for each time course. To adjust for multiple testing, we applied the Benjamini & Hochberg method to the one-tailed p-values corresponding to the observed z-scores. This allowed us to establish a z-score threshold based on the equivalent corrected p-value. Only Z-scores profiles with Z-scores greater than the threshold at more than two consecutive times over the day were retained in the analysis. The hypothesis for filtering Z-score profiles in this last step is that if an element were to be active based on its overrepresentation, then it would be overrepresented in adjacent phases of the day. Then the 3–8mer was assigned a phase value (time in hrs from lights on or subjective dawn) based on the time of day that adjacent Z-scores are significant. The phase Z-score profile phase was then used to cluster similar Z-score profiles, and compare between species.

Papaya Sequence Used In This Study

Carica genome sequence, promoter sequence and predicted proteins can be found at our web site: http://diurnal-files.cgrb.oregonstate.edu/papaya_sequence/

Acknowledgements We thank Kerry Lutz for comments on this work. This work was supported by Waksman Startup funds to T.P.M, and partially supported by NSF Plant Genome Grant DBI 0605240 to T.C.M. H.D.P was supported by a Computational and Genome Biology Initiative Fellowship from Oregon State University. F.A. was supported by Waksman Undergraduate Research Fellowship, and W.W. was supported by a Rutgers Energy Institute graduate fellowship.

References

1. Alabadí D, Oyama T, Yanovsky MJ, Harmon FG, Más P, Kay SA (2001) Reciprocal regulation between *TOC1* and *LHY/CCA1* within the *Arabidopsis* circadian clock. Science 293:880–883. doi:10.1126/science.1061320
2. Blasing O, Gibon Y, Gunther M, Hohne M, Morcuende R, Osuna D, Thimm O, Usadel B, Scheible W, Stitt M (2005) Sugars and circadian regulation make major contributions to the global regulation of diurnal gene expression in *Arabidopsis*. Plant Cell 17:3257–3281. doi:10.1105/tpc.105.035261
3. Borchert R, Renner SS, Calle Z, Navarrete D, Tye A, Gautier L, Spichiger R, von Hildebrand P (2005) Photoperiodic induction of synchronous flowering near the Equator. Nature 433:627–629. doi:10.1038/nature03259
4. Boxall SF, Foster JM, Bohnert HJ, Cushman JC, Nimmo HG, Hartwell J (2005) Conservation and divergence of circadian clock operation in a stress-inducible crassulacean acid metabolism species reveals clock compensation against stress. Plant Physiol 137:969–982. doi:10.1104/pp.104.054577
5. Covington M, Maloof J, Straume M, Kay S, Harmer S (2008) Global transcriptome analysis reveals circadian regulation of key pathways in plant growth and development. Genome Biol 9:R130. doi:10.1186/gb-2008-9-8-r130
6. Dodd AN, Salathia N, Hall A, Kevei E, Toth R, Nagy F, Hibberd JM, Millar AJ, Webb AA (2005) Plant circadian clocks increase photosynthesis, growth, survival, and competitive advantage. Science 309:630–633. doi:10.1126/science.1115581
7. Edwards KD, Anderson PE, Hall A, Salathia NS, Locke JC, Lynn JR, Straume M, Smith JQ, Millar AJ (2006) *FLOWERING LOCUS C* mediates natural variation in the high-temperature response of the *Arabidopsis* circadian clock. Plant Cell 18:639–650. doi:10.1105/tpc.105.038315
8. Facella P, Lopez L, Carbone F, Galbraith DW, Giuliano G, Perrotta G (2008) Diurnal and circadian rhythms in the tomato transcriptome and their modulation by Cryptochrome photoreceptors. PLoS One 3:e2798. doi:10.1371/journal.pone.0002798
9. Farre E, Harmer S, Harmon F, Yanovsky M, Kay S (2005) Overlapping and distinct roles of PRR7 and PRR9 in the Arabidopsis circadian clock. Curr Biol 15(1):47–54. doi:10.1016/j.cub.2004.12.067
10. Farre EM, Kay SA (2007) PRR7 protein levels are regulated by light and the circadian clock in *Arabidopsis*. Plant J 52:548–560. doi:10.1111/j.1365-313X.2007.03258.x
11. Fujiwara S, Wang L, Han L, Suh S-S, Salome PA, McClung CR, Somers DE (2008) Post-translational regulation of the *Arabidopsis* circadian clock through selective proteolysis and phosphorylation of Pseudo-response Regulator proteins. J Biol Chem 283:23073–23083. doi:10.1074/jbc.M803471200
12. Harmer S, Kay S (2005) Positive and negative factors confer phase-specific circadian regulation of transcription in *Arabidopsis*. Plant Cell 17:1926–1940. doi:10.1105/tpc.105.033035
13. Hazen SP, Schultz TF, Pruneda-Paz JL, Borevitz JO, Ecker JR, Kay SA (2005) *LUX ARRHYTHMO* encodes a Myb domain protein essential for circadian rhythms. Proc Natl Acad Sci USA 102:10387–10392. doi:10.1073/pnas.0503029102
14. Hecht V, Knowles CL, Vander Schoor JK, Liew LC, Jones SE, Lambert MJM, Weller JL (2007) Pea LATE BLOOMER1 Is a GIGANTEA ortholog with roles in photoperiodic flowering, deetiolation, and transcriptional regulation of circadian clock gene homologs. Plant Physiol 144:648–661. doi:10.1104/pp.107.096818
15. Kiba T, Henriques R, Sakakibara H, Chua N-H (2007) Targeted degradation of PSEUDO-RESPONSE REGULATOR5 by an SCFZTL complex regulates clock function and photomorphogen-

esis in *Arabidopsis thaliana*. Plant Cell 19:2516–2530. doi:10.1105/tpc.107.053033
16. Kim W-Y, Geng R, Somers DE (2003) Circadian phase-specific degradation of the F-box protein ZTL is mediated by the proteasome. Proc Natl Acad Sci USA 100:4933–4938. doi:10.1073/pnas.0736949100
17. Kim W-Y, Fujiwara S, Suh S-S, Kim J, Kim Y, Han L, David K, Putterill J, Nam HG, Somers DE (2007) ZEITLUPE is a circadian photoreceptor stabilized by GIGANTEA in blue light. Nature 449:356. doi:10.1038/nature06132
18. Liu H, Wang H, Gao P, X, J, X, T, Wang J, Wang B, Lin C, Fu Y-F (2008) Analysis of clock gene homologs using unifoliolates as target organs in soybean (*Glycine max*). J Plant Physiol (in press).
19. Locke J, Kozma-Bognar L, Gould P, Feher B, Kevei E, Nagy F, Turner M, Hall A, Millar A (2006) Experimental validation of a predicted feedback loop in the multi-oscillator clock of *Arabidopsis thaliana*. Mol Syst Biol 2:59. doi:10.1038/msb4100102
20. Matsuo T, Okamoto K, Onai K, Niwa Y, Shimogawara K, Ishiura M (2008) A systematic forward genetic analysis identified components of the *Chlamydomonas* circadian system. Genes Dev 22:918–930. doi:10.1101/gad.1650408
21. McClung C (2006) Plant circadian rhythms. Plant Cell 18(4):792–803. doi:10.1105/tpc.106.040980
22. Michael T, Mockler T, Breton G, McEntee C, Byer A, Trout J, Hazen S, Shen R, Priest H, Sullivan C, Givan S, Yanovsky M, Hong F, Kay S, Chory J (2008a) Network discovery pipeline elucidates conserved time of day specific cis-regulatory modules. PLoS Genet 4:e14. doi:10.1371/journal.pgen.0040014
23. Michael TP, McClung CR (2002) Phase-specific circadian clock regulatory elements in *Arabidopsis*. Plant Physiol 130:627–638. doi:10.1104/pp.004929
24. Michael TP, McClung CR (2003) Enhancer trapping reveals widespread circadian clock transcriptional control in *Arabidopsis thaliana*. Plant Physiol 132:629–639. doi:10.1104/pp.021006
25. Michael TP, Salomé PA, McClung CR (2003a) Two *Arabidopsis* circadian oscillators can be distinguished by differential temperature sensitivity. Proc Natl Acad Sci USA 100:6878–6883. doi:10.1073/pnas.1131995100
26. Michael TP, Breton G, Hazen SP, Priest H, Mockler TC, Kay SA, Chory J (2008b) A morning-specific phytohormone gene expression program underlying rhythmic plant growth. PLoS Biol 6:e225. doi:10.1371/journal.pbio.0060225
27. Michael TP, Salome PA, Yu HJ, Spencer TR, Sharp EL, McPeek MA, Alonso JM, Ecker JR, McClung CR (2003b) Enhanced fitness conferred by naturally occurring variation in the circadian clock. Science 302:1049–1053. doi:10.1126/science.1082971
28. Ming R et al (2008) The draft genome of the transgenic tropical fruit tree papaya (Carica papaya Linnaeus). Nature 452:991–996. doi:10.1038/nature06856
29. Miwa K, Serikawa M, Suzuki S, Kondo T, Oyama T (2006) Conserved expression profiles of circadian clock-related genes in two Lemna species showing long-day and short-day photoperiodic flowering responses. Plant Cell Physiol 47:601–612. doi:10.1093/pcp/pcj027
30. Mockler TC, Michael TP, Priest HD, Shen R, Sullivan CM, Givan SA, McEntee C, Kay S, Chory J (2007) THE DIURNAL PROJECT: Diurnal and circadian expression profiling, model-based pattern matching and promoter analysis. Cold Spring Harbor Symposia on Quantitative Biology: Clocks and Rhythms 72.
31. Mockler TC, Yu X, Shalitin D, Parikh D, Michael TP, Liou J, Huang J, Smith Z, Alonso JM, Ecker JR, Chory J, Lin C (2004) Regulation of flowering time in *Arabidopsis* by K homology domain proteins. Proc Natl Acad Sci USA 101:12759–12764. doi:10.1073/pnas.0404552101
32. Murakami M, Ashikari M, Miura K, Yamashino T, Mizuno T (2003) The evolutionarily conserved OsPRR quintet: rice pseudo-response regulators implicated in circadian rhythm. Plant Cell Physiol 44:1229–1236. doi:10.1093/pcp/pcg135
33. Murakami M, Matsushika A, Ashikari M, Yamashino T, Mizuno T (2005) Circadian-associated rice pseudo response regulators (OsPRRs): insight into the control of flowering time. Biosci Biotechnol Biochem 69:410–414. doi:10.1271/bbb.69.410
34. Ramos A, Perez-Solis E, Ibanez C, Casado R, Collada C, Gomez L, Aragoncillo C, Allona I (2005) Winter disruption of the circadian clock in chestnut. Proc Natl Acad Sci USA 102:7037–7042. doi:10.1073/pnas.0408549102
35. Schaffer R, Ramsay N, Samach A, Corden S, Putterill J, Carré IA, Coupland G (1998) *LATE ELONGATED HYPOCOTYL*, an Arabidopsis gene encoding a MYB transcription factor, regulates circadian rhythmicity and photoperiodic responses. Cell 93:1219–1229. doi:10.1016/S0092-8674(00)81465-8
36. Somers DE, Schultz TF, Milnamow M, Kay SA (2000) ZEITLUPE encodes a novel clock-associated PAS protein from *Arabidopsis*. Cell 101:319–329. doi:10.1016/S0092-8674(00)80841-7
37. Strayer C, Oyama T, Schultz TF, Raman R, Somers DE, Más P, Panda S, Kreps JA, Kay SA (2000) Cloning of the *Arabidopsis* clock gene *TOC1*, an autoregulatory response regulator homolog. Science 289:768–771. doi:10.1126/science.289.5480.768
38. Turner A, Beales J, Faure S, Dunford RP, Laurie DA (2005) The pseudo-response regulator Ppd-H1 provides adaptation to photoperiod in barley. Science 310:1031–1034. doi:10.1126/science.1117619
39. Wang Z-Y, Tobin EM (1998) Constitutive expression of the *CIRCADIAN CLOCK ASSOCIATED 1* (*CCA1*) gene disrupts circadian rhythms and suppresses its own expression. Cell 93:1207–1217. doi:10.1016/S0092-8674(00)81464-6
40. Zeilinger M, Farre E, Taylor S, Kay S, Doyle F (2006) A novel computational model of the circadian clock in *Arabidopsis* that incorporates *PRR7* and *PRR9*. Mol Syst Biol 2:58. doi:10.1038/msb4100101