Evolutionarily Conserved Regulatory Motifs in the Promoter of the *Arabidopsis* Clock Gene *LATE ELONGATED HYPOCOTYL* [C][W]

Mark Spensley,[a,1] Jae-Yean Kim,[a,2] Emma Picot,[a,b] John Reid,[c] Sascha Ott,[b] Chris Helliwell,[d] and Isabelle A. Carré[a,3]

[a] Department of Biological Sciences, University of Warwick, Coventry CV4 7AL, United Kingdom
[b] Systems Biology Centre, University of Warwick, Coventry CV4 7AL, United Kingdom
[c] MRC Biostatistics Unit, Institute of Public Health, University Forvie Site, Cambridge CB2 0SR, United Kingdom
[d] CSIRO Plant Industry, Canberra, ACT 2601, Australia

The transcriptional regulation of the *LATE ELONGATED HYPOCOTYL* (*LHY*) gene is key to the structure of the circadian oscillator, integrating information from multiple regulatory pathways. We identified a minimal region of the *LHY* promoter that was sufficient for rhythmic expression. Another upstream sequence was also required for appropriate waveform of transcription and for maximum amplitude of oscillations under both diurnal and free-running conditions. We showed that two classes of protein complexes interact with a G-box and with novel 5A motifs; mutation of these sites reduced the amplitude of oscillation and broadened the peak of expression. A genome-wide bioinformatic analysis showed that these sites were enriched in phase-specific clusters of rhythmically expressed genes. Comparative genomic analyses showed that these motifs were conserved in orthologous promoters from several species. A position-specific scoring matrix for the 5A sites suggested similarity to CArG boxes, which are recognized by MADS box transcription factors. In support of this, the FLOWERING LOCUS C (FLC) protein was shown to interact with the *LHY* promoter in planta. This suggests a mechanism by which FLC might affect circadian period.

INTRODUCTION

The circadian clock enables plants to adapt their physiology in anticipation of predictable daily changes in light and temperature conditions (Harmer, 2009). Correct matching of the clock's endogenous period with environmental day–night cycles has been shown to confer a fitness advantage (Dodd et al., 2005). This fitness advantage is thought to reflect the appropriate timing of circadian outputs in relation to dawn and dusk. For example, many components of metabolic pathways are under circadian control, as are genes controlling growth or responses to biotic and abiotic stress (Harmer et al., 2000). Optimal timing of these activities relative to environmental cycles is likely to contribute to the amount of biomass produced. Furthermore, seasonal responses also rely on the appropriate timing of gene expression rhythms, since the photoperiodic induction of flowering in

Arabidopsis thaliana is triggered when the circadian rhythm of CONSTANS gene expression coincides with light under long-day conditions (Suarez-Lopez et al., 2001; Roden et al., 2002; Yanovsky and Kay, 2002). Thus, elucidating the mechanism of the clock and understanding the factors that determine the precise timing of downstream rhythms will open up new avenues for crop improvement.

A large portion of the genome is under circadian control, suggesting that transcriptional regulation forms the root of many circadian output pathways. Up to 89% of the genome has been shown to exhibit rhythmic expression under at least one experimental condition (Michael et al., 2008). However, not much is known about the transcription factors that mediate rhythmic transcription and how they interact to generate specific phases and waveforms of transcription. Here, we used a combination of experimental and bioinformatic approaches to identify regulatory elements that mediate circadian transcription of the *LATE ELONGATED HYPOCOTYL* (*LHY*) gene in *Arabidopsis*. *LHY* encodes a MYB transcription factor that functions redundantly with *CIRCADIAN CLOCK ASSOCIATED1* (*CCA1*) at the core of the circadian oscillator (Schaffer et al., 1998; Alabadi et al., 2001, 2002; Mizoguchi et al., 2002). Current models place *LHY* and *CCA1* at the intersection of either two or three regulatory feedback loops involving *TIMING of CAB1* (*TOC1*) and *PSEUDO-RESPONSE REGULATOR7* (*PRR7*) and *PRR9*, respectively (Locke et al., 2006; Zeilinger et al., 2006). Thus, *LHY* and *CCA1* occupy a central position within the circadian network. Their transcription is also regulated by light, a feature that is important for entrainment of the circadian clock to light–dark cycles. We

therefore reasoned that analysis of the transcriptional regulation of *LHY* should reveal aspects of the network structure by identifying rhythmic inputs from different oscillators. At the same time, this study would uncover the logic of interactions between circadian-regulated and light-regulated promoter elements, which ultimately determines the precise timing of transcription.

Transcription factor binding sites could in theory be identified by searching promoter sequences for matches to known position-specific scoring matrices (PSSMs) found in databases. However, the information available for plants is limited at this stage. Approximately 150 such matrices are currently available, which are clearly insufficient to account for >2000 transcription factors encoded in the *Arabidopsis* genome (Riechmann et al., 2000; Guo et al., 2005). In silico discovery of binding sites is further hampered by a high false positive rate. Several promoter elements have been associated with circadian regulation so far. For example, the CCA1 binding site (AAAAATCT) was found in the promoter of midday-specific *lhcb* genes, encoding light-harvesting chlorophyll a/b binding proteins (Carré and Kay, 1995; Wang et al., 1997). CCA1 and LHY also bind a related sequence named the evening element (AAATATCT), which is overrepresented in sets of evening-specific promoters (Harmer et al., 2000). Both CCA1 binding site and EE elements were shown to specify circadian phase and to be sufficient for rhythmic transcription (Michael and McClung, 2002; Harmer and Kay, 2005). The G-box core sequence (CACGTG), which has a well-characterized role in mediating responses to light, and the related Hex element (TGACGTGG) were found to be overrepresented in the promoters of clock-regulated genes and to be enriched in sets of dawn-specific genes with the consensus GACACGTGG (Michael and McClung, 2003; Michael et al., 2008), but the role of these sequences in conferring phase-specific expression is less well established. A motif described as the morning element (AACCAC) was found to confer morning-specific expression to the *PRR9* promoter (Harmer and Kay, 2005). This sequence was related to a Sequence Over-Represented in Light-Induced Promoters (SORLIP 1; GCCAC) and overlapped with a sequence enriched in the promoters of clock-regulated genes (CACT<u>AACCAC</u>) (Hudson and Quail, 2003). A more refined consensus sequence for the morning element (CCACAC) was obtained through analysis of a large microarray data set and shown to be associated with morning-specific gene expression (Michael et al., 2008). Other motifs that show time of the day–specific enrichment in rhythmic promoters include an evening-specific GATA element (GGATAAG) and the late night–specific telo box (AAACCCT), starch box (AAGCCC), and protein box (ATGGGCC) (Michael et al., 2008). A functional genomics approach recently showed that the transcription factor CCA1 HIKING EXPEDITION (CHE) functions as a rhythmic repressor of *CCA1* expression. Loss of CHE function or disruption of its binding site (GGNCCCAC) did not abolish rhythmic transcription from the *CCA1* promoter, suggesting the contribution of one or more additional rhythmic signals (Pruneda-Paz et al., 2009). The *LHY* promoter does not contain CHE binding sites, and the mechanisms underlying its rhythmic activity remain unclear.

In this article, we experimentally define regions of the *LHY* promoter that mediate circadian regulation. Within these regions, we identify sites that are bound by protein complexes in vitro and affect circadian expression in vivo. We investigate the function of two types of promoter motifs, including a G-box and a CArG-like sequence described as the 5A motif. We further perform a statistical analysis of the genome-wide function of these motifs in the control of rhythmic gene expression. Using a comparative genomics technique, we detect an evolutionarily conserved region in the *LHY* promoter that matches the region defined experimentally. We find the G-box and 5A motifs to be conserved, therefore providing further evidence for their functional importance. In addition, we unravel conserved sequence patterns in the *LHY* promoter that are also present in the *CCA1* promoter and provide promising targets for future experiments. Furthermore, we demonstrate binding of the MADS box protein FLOWERING LOCUS C (FLC) to the *LHY* promoter, suggesting a mechanism by which this transcription factor might modulate circadian period to ensure its temperature compensation.

RESULTS

Mapping of 5′ Upstream Sequences Controlling the Rhythmic Expression of *LHY*

Previous results showed that a reporter construct (−*1618 PLHY: luc*, previously described as *PLHY:luc1*), consisting of 1618 bp upstream of the translational start site of *LHY* fused to a *luciferase* (*luc*) reporter gene and to the 3′ untranslated region (UTR) of the *nopaline synthase gene* (*nos*) fully recapitulated the rhythmic pattern of expression of the endogenous *LHY* transcript (Kim et al., 2003). The full 5′ upstream sequence of this construct is given in Supplemental Figure 1 online. To further delimit the upstream regulatory region of *LHY*, a set of 5′ deletion constructs was generated (Figure 1A). Rhythmic expression was analyzed in transgenic plants, first under diurnal light–dark (LD) cycles (where expression patterns reflect dual control by light and by the circadian clock), then upon transfer to constant light (LL; where expression patterns strictly reflect control by the circadian clock). Both blue and red light conditions were tested, since the contrasting phenotypes of TOC1-RNA interference lines under these conditions suggested that the circadian regulatory network might operate differently under these different light qualities (Mas et al., 2003). As similar expression patterns were observed, we only show results for red light experiments.

Under 12L12D cycles, expression of the full-length −*1618 PLHY:luc* construct began to rise ∼4 h before dawn (Figure 1B). A sharp increase in luminescence was observed in response to the light-on signal, a peak was reached 2 to 4 h later, and photon counts returned to trough levels in the evening. Rhythmicity persisted following transfer to constant light, but with reduced amplitude (Figure 1B).

Deletion of 5′ sequences of the *LHY* promoter to 957 bp upstream of the translational start site (−957*PLHY:luc* construct) did not alter expression levels nor the pattern of rhythmic expression under either LD or LL (see Supplemental Figure 2 online; Figure 1B). A further deletion to position −847 (−847 *PLHY:luc* construct) reduced the amplitude of the luminescence rhythm under both LD and LL (Figure 1B, Table 1). The onset of
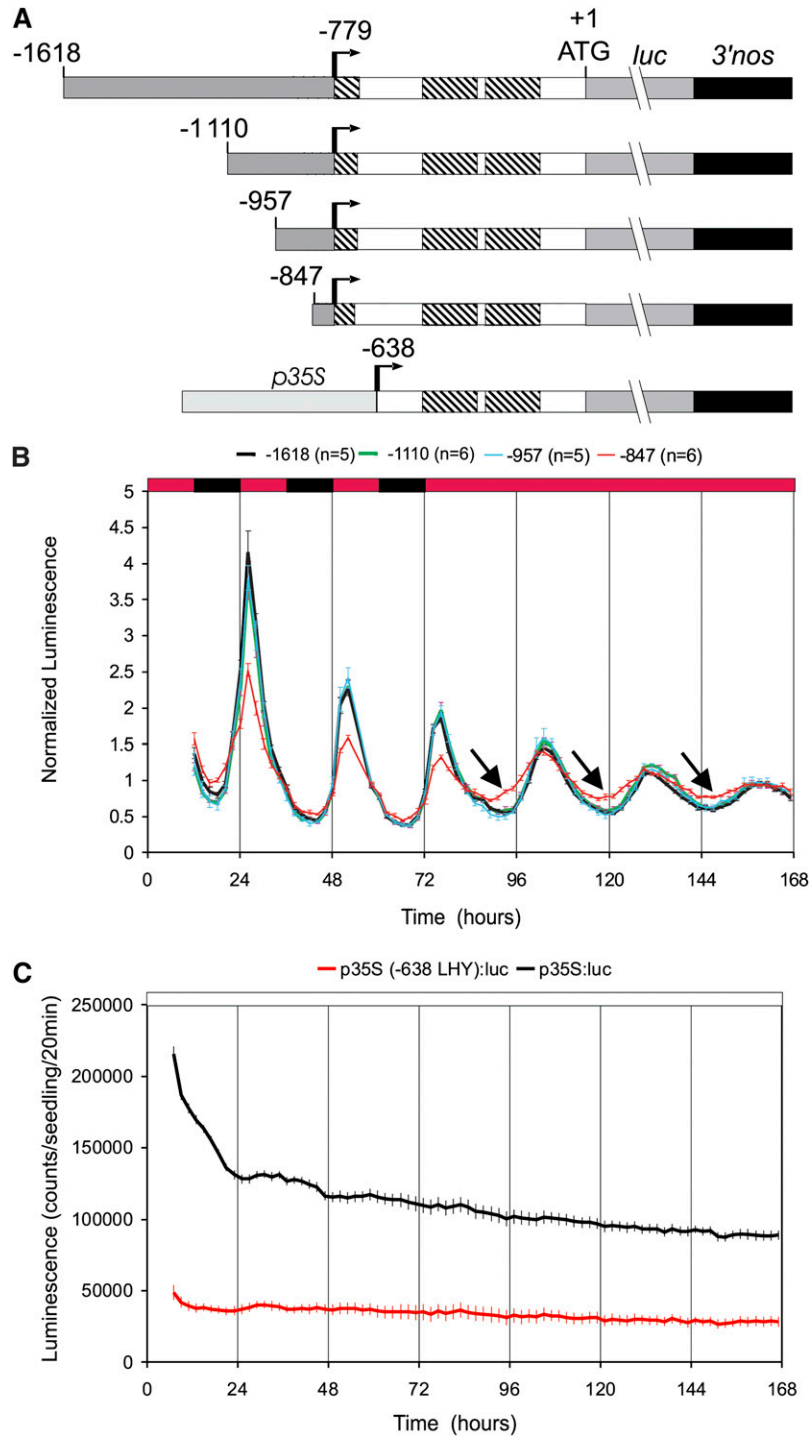
Figure 1. 5′ Deletion Analysis of the *LHY* Upstream Region.

(A) The 5′ upstream sequence of *LHY* (starting either 1618, 1110, 957, 847, or 638 bp upstream of the translational start site) was fused to the translational start site of firefly *luciferase* and to the 3′ UTR of the *nos* gene. The fragment starting at position −638 lacked the transcriptional start site (indicated at position −779) and was placed downstream of the cauliflower mosaic virus *35S* promoter. Hatched boxes within the 5′ UTR indicate introns (not drawn to scale).

(B) Rhythmic expression patterns of 7-d-old transgenic plants placed under 12L12D cycles of red light and then transferred to constant light at time 72 h. Red and dark boxes at the top of the graphs indicate intervals of red light or darkness, respectively. Arrows highlight the early onset of transcription for the −*847PLHY:luc* construct.

Table 1. FFT-NLLS Analysis of *PLHY:luc* Expression Patterns in Constant Light

pLHY:luc Construct	Photoperiod during Entrainment	Period (h)	Amplitude	Phase (h)[a]	Skewness	Kurtosis	RAE[b]	*n*
−957	4L 20D	27.20 ± 0.22	0.24 ± 0.02	−0.70 ± 0.65	0.29 ± 0.05	2.71 ± 0.10	0.29 ± 0.01	12
	8L16D	28.86 ± 0.16	0.60 ± 0.04	−0.73 ± 0.42	0.24 ± 0.03	2.69 ± 0.07	0.32 ± 0.02	16
	12L12D	28.41 ± 0.18	0.45 ± 0.06	−1.69 ± 0.41	0.21 ± 0.04	2.48 ± 0.05	0.34 ± 0.01	22
	16L8D	29.11 ± 0.21	0.33 ± 0.04	+0.49 ± 0.37	0.08 ± 0.14	2.04 ± 0.07	0.37 ± 0.05	8
	20L8D	26.16 ± 0.12	0.25 ± 0.03	−5.15 ± 2.99	−0.04 ± 0.04	1.95 ± 0.08	0.22 ± 0.05	8
−957 class IA	8L16D	29.25 ± 0.14	0.37 ± 0.01[###]	−2.63 ± 0.17[##]	0.14 ± 0.03[#]	2.64 ± 0.05	0.29 ± 0.01	10
−957 class IB	8L16D	28.95 ± 0.12	0.33 ± 0.01[###]	−2.93 ± 0.18[###]	0.15 ± 0.03[#]	2.58 ± 0.04	0.33 ± 0.01	12
−957 class IIA	8L16D	29.12 ± 0.16	0.25 ± 0.01[###]	−2.44 ± 0.14[#]	0.06 ± 0.05[##]	2.46 ± 0.08	0.32 ± 0.01	6
−957 class IIB	8L16D	28.81 ± 0.11	0.39 ± 0.03[###]	−0.80 ± 0.52	0.16 ± 0.02[#]	2.55 ± 0.03	0.34 ± 0.01	18
−957 1,2m	8L16D	28.91 ± 0.11	0.41 ± 0.03	−1.38 ± 0.54	0.19 ± 0.02	2.68 ± 0.04	0.40 ± 0.02	18
	12L12D	29.58 ± 0.18	0.40 ± 0.04	+0.01 ± 0.60[#]	−0.016 ± 0.09[#]	2.66 ± 0.02	0.30 ± 0.01	12
−957 3,4,5m	12L12D	28.84 ± 0.10	0.29 ± 0.03	−0.14 ± 1.20	0.20 ± 0.05	2.39 ± 0.09	0.29 ± 0.02	8
−847	4L20D	27.22 ± 0.25	0.18 ± 0.01**	−1.38 ± 0.29	0.09 ± 0.03***	2.49 ± 0.05*	0.34 ± 0.02	17
	8L16D	28.89 ± 0.25	0.38 ± 0.01***	+1.10 ± 0.36**	0.15 ± 0.01**	2.56 ± 0.02*	0.33 ± 0.02	22
	12L12D	28.26 ± 0.24	0.32 ± 0.03*	−0.97 ± 0.47	0.07 ± 0.05*	2.41 ± 0.06	0.30 ± 0.02	29
	16L8D	29.50 ± 0.25	0.21 ± 0.02**	+1.11 ± 0.50	0.03 ± 0.06	2.46 ± 0.18	0.37 ± 0.03	12
	20L8D	25.69 ± 0.33	0.11 ± 0.02***	−2.58 ± 2.97	0.08 ± 0.06	2.16 ± 0.17	0.46 ± 0.06***	9
−847 3,4,5 m	8L16D	28.75 ± 0.30	0.32 ± 0.02[#]	+2.09 ± 0.40	0.14 ± 0.03	2.60 ± 0.05	0.37 ± 0.02	6
	12L12D	28.79 ± 0.08	0.26 ± 0.02	+1.26 ± 0.33[##]	0.19 ± 0.06	2.34 ± 0.03	0.28 ± 0.01	12
	16L8D	28.19 ± 0.43	0.22 ± 0.4	−1.92 ± 0.80[##]	0.00 ± 0.11	2.53 ± 0.13	0.42 ± 0.02	6

FFT-NLLS analysis was carried out between 24 and 130 h in constant light on data that had been normalized to average expression levels in constant light. At least three independent transgenic lines were tested for each construct. Results for each condition were pooled for at least two independent experiments. *, **, and *** indicate P values of 0.05, 0.01, and 0.001 for differences between truncated (*−847 PLHY:luc*) and full-length (*−957 PLHY:luc*) constructs. [#], [##], and [###] indicate P values of 0.05, 0.01, and 0.001 for differences between mutant constructs and the corresponding wild-type construct (*−847* or *−957 PLHY:luc*).
[a]Peak phase relative to dawn. + indicates a phase advance; − indicates a phase delay.
[b]Relative amplitude error (RAE) values are indicative of the quality of the fit to a cosine wave, with RAE values closer to 1 indicative of weaker rhythms.

luciferase expression was notably advanced in LL (Figure 1B, arrows). This change in waveform was also observed under white or blue light conditions (see Supplemental Figures 3A and 3B online) and was consistent across all six transgenic lines tested (see Supplemental Figure 3C online).

The *−847 PLHY:luc* construct comprised 128 bp upstream of the transcriptional start site and 719 bp of 5′ UTR. As this 5′ UTR sequence comprises three introns and several short open reading frames that may play a role in translational regulation of *LHY* expression, we questioned whether circadian regulation of *−847 PLHY:luc* expression might take place at the posttranscriptional level. RNA blot analysis of transgene expression indicated that the *luc* mRNA accumulated rhythmically with a level and amplitude similar to that of the endogenous *LHY* transcript (see Supplemental Figure 4 online). No changes in transcript size were detected in this experiment, making regulation by differential splicing highly unlikely. These results indicate that rhythmic expression of *−847 PLHY:luc* expression is controlled either at the transcriptional level or at the level of mRNA stability. However, this doesn't preclude additional levels of regulation at the translational level. For example, our previous work showed that translation of the *LHY* mRNA was upregulated in response to light signals (Kim et al., 2003).

Regulation of mRNA stability is mainly mediated by *cis*-acting elements located within the 3′ UTR of the mRNA (Mignone et al., 2002). To test the potential contribution of the 3′ UTR to *LHY* expression patterns, the *nos* 3′ UTR of *−1618 PLHY:luc* was substituted for the *LHY* 3′ UTR to give *PLHY:luc3*. This construct did not have significantly altered timing of luciferase expression in transgenic plants, whether under LD cycles or in constant conditions (see Supplemental Figure 5 online). Similarly, replacement of most of the *LHY* 5′ UTR (from position −638 to the ATG) with the *nos* 5′ UTR did not alter the temporal pattern of luminescence. Therefore, 5′ and 3′ UTR sequences were not

Figure 1. (continued).

(C) Arrhythmic expression of the *p35S (−638 LHY):luc* construct in constant light (from mixed red and blue LEDs). Expression of a *p35S:luc* construct is shown as a control. Plants were grown for 7 d under LD cycles and then transferred to constant light at time zero. Each of the data points in **(B)** and **(C)** represents average expression levels for *n* independent transgenic lines, normalized relative to mean expression levels in constant light. Error bars indicate SE. All experiments were performed at least twice with similar results.

essential for the rhythmic expression pattern of *LHY*. Moreover, sequences from position −638 to +1 of the LHY gene did not confer rhythmic expression when inserted downstream of the *35S* promoter of the cauliflower mosaic virus *P35S(-618LHY):luc* construct (Figure 1C), showing that 5′ UTR sequences of *LHY* were not sufficient to mediate circadian regulation.

Altogether, these results indicated that sequences mediating rhythmic transcription of *LHY* were located in a 210-bp region between 847 and 638 bases upstream of the translational start site.

Previous work showed that expression of the endogenous *LHY* transcript was repressed in trangenic plants that carried an overexpressed copy of the gene (Schaffer et al., 1998). This provided evidence that *LHY* functions as part of a negative transcriptional feedback loop. Supplemental Figure 6 online shows that expression of the *847PLHY:luc* construct is reduced to wild-type trough levels in *LHY-ox* plants. Therefore the −847/+1 region of the *LHY* promoter also contains a regulatory element(s) mediating negative autoregulation.

The −957/−847 Region of the LHY Promoter Mediates Photoperiod-Dependent Changes in Transcriptional Waveform

The results above suggested that the −957/−847 region of the *LHY* promoter acts to delay the onset of transcription. To further characterize the function of this promoter fragment, expression of the −957 and −847 *PLHY:luc* constructs was compared in plants that were entrained to different photoperiods then transferred to constant light. Under 8L16D, expression of −847 *PLHY: luc* began to rise 6 h earlier than that of −957 *PLHY:luc* (Figure 2A). By contrast, no significant difference was observed under 16L8D (Figure 2B). Figure 2C shows that the trough of expression for the truncated construct was advanced by an average of 2 to 4 h under photoperiods 12 h and under, but not under longer photoperiods. Fast Fourier transform-nonlinear least square (FFT-NLLS; Plautz et al., 1997) analysis of the data (shown in Table 1) detected a significant phase advance for the −847
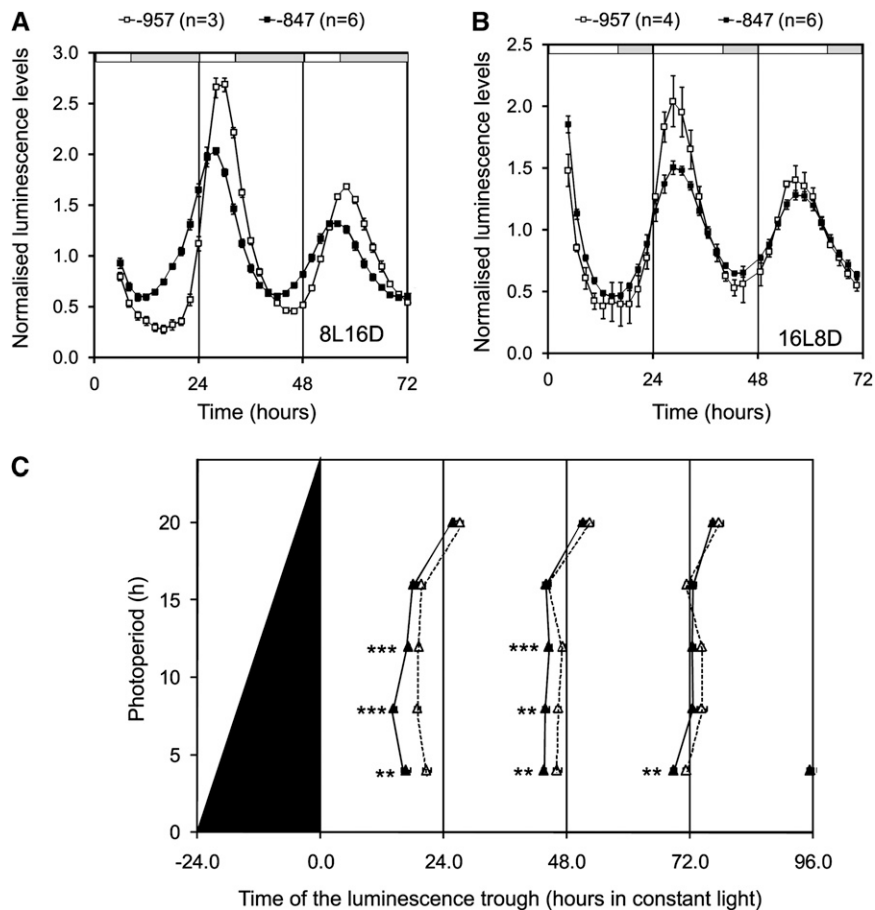

Figure 2. Differential Phase Adjustment of the −957 and −847PLHY:luc Constructs in Response to Changing Photoperiods.

(A) and **(B)** Rhythmic luminescence patterns from plants grown under cycles of either 8L16D or 16L8D of white light for 7 d and then subjected to a further three photocycles of red light before release into constant red light at time zero.

(C) Similar experiments were performed for photoperiods ranging from 4L20D to 20L4D. The black wedge indicates the last interval of darkness prior to transfer to constant light. The times of the troughs of luminescence are indicated by closed and open triangles for the −847 and −957PLHY:luc constructs, respectively. Each data point represents data averaged from three to six transgenic lines and at least three independent experiments. Significant differences are indicated by * (P < 0.05), ** (P < 0.01), or *** (P < 0.001).

PLHY:luc construct following entrainment to 8L16D but not to other photoperiods. As FFT-NLLS fits a cosine wave to the data, which may not detect features of complex oscillations, we also performed a waveform analysis. This returned positive skewness values for the −957 *PLHY:luc* construct under photoperiods 12 h or under, indicating that the peak of expression was asymmetric with a faster rise and a slower decay. This asymmetry was not detected with the −847 *PLHY:luc* construct under any photoperiod.

Altogether, these observations showed that the −957/−847 fragment of the *LHY* promoter contains an element that represses transcription in the late subjective night to delay its onset until subjective dawn. The photoperiod dependency of this effect may be explained by the latest models for the *Arabidopsis* circadian clock, which place *LHY* at the convergence of multiple oscillatory feedback loops (Locke et al., 2006; Zeilinger et al., 2006). Our results suggest that the waveform of *LHY* transcription reflects the composite action of rhythmic transcriptional activators as well as repressors. If these different activities mediate signals from distinct oscillators, any photoperiod-driven change in the phase relationship between these different oscillators would be expected to result in alterations of *LHY* expression waveform. For example, under short-day conditions, the effect of a transcriptional repressor may overlap with the effect of a transcriptional activator, delaying the onset of transcription and resulting in asymmetric peaks. Under long-day conditions, the transcriptional repressor may oscillate out of phase with the transcriptional activator and therefore have no effect on the onset of transcription. This would explain the change to a symmetrical waveform of transcription. In this model, deletion of the −957/−847 region would have disrupted the effect of the hypothetical repressor, resulting in a symmetrical waveform of transcription under all photoperiods. Further work will be required to probe this hypothesis.

Mapping of Putative Transcription Factor Binding Sites

To identify transcription factor complexes that might contribute to the rhythmic expression pattern of *LHY*, electrophoretic mobility shift assays (EMSAs) were performed using whole-cell plant extracts and radiolabeled fragments of the *LHY* promoter (Figures 3B and 3E; quantification of EMSAs is shown in Supplemental Figure 7 online).

The −957 to −847 fragment identified four groups of DNA–protein complexes (marked I to IV in Figure 3B). The position of binding sites within the probe was then narrowed down by competition assays using an array of overlapping 30-bp promoter fragments. Binding of group I was severely reduced in the presence of 100-fold molar excess of competitor 6. This oligonucleotide was centered on a G-box sequence (CACGTG), which is known to play a role in the light regulation of *LHY* transcription (Martinez-Garcia et al., 2000). Competitor 7 was less effective, even though it also contained the G-box. Thus, sequences flanking the G-box were also important for DNA binding. Complex IV was outcompeted by oligonucleotides 3, 4, 8, and 9. Complexes II and III showed overlapping specificity with complex IV, suggesting that they may correspond to the same protein binding to multiple sites on the promoter. Different bands

may arise from different numbers of transcription factor molecules binding to the DNA or from association with different cofactors. Visual inspection of the most effective competitors identified the common sequence AAAAA (or TTTTT in reverse orientation); therefore, we named this putative binding site the 5A motif. Interestingly, binding of complex I was reduced in the presence of competitor oligonucleotides 4 and 8 comprising the 5A motif, and this suggested a possible interaction between group I and II complexes.

The −847 to −757 fragment of the *LHY* promoter identified two complexes, labeled V and VI in Figure 3E. Binding sites were mapped as above, by competition assays using an array of overlapping 30-bp oligonucleotides. Both complexes were outcompeted by oligonucleotide 14, and to a lesser extent, 15. Competitor 14 comprised three copies of the 5A motif, one of which was in common with the overlapping competitor 15. This suggested that the −847/−757 sequence might contain further binding sites for the biochemical activity identified using the −957/−847 probe. To test this hypothesis, we tested whether protein complexes forming on oligonucleotide 14 also showed affinity for oligonucleotides 9 and 3. Supplemental Figure 8 online shows that oligonucleotides 9 and 3 were equally capable of competing for binding to oligonucleotide 14 as excess unlabeled probe. Altogether, these results suggested that the transcription factor interacting with the 5A motif had two binding sites within the −957 to −847 region of the *LHY* promoter and up to three additional binding sites between positions −847 and −757 (Figure 3G).

In Vivo Effects of G-Box Mutations

To test the role of the G-box sequence and 5A motifs in the regulation of *LHY* gene expression, we identified point mutations that disrupted protein binding to these sequences in vitro and then tested for effects of these mutations on expression of our *PLHY:luc* reporter constructs in vivo. Mutation of the core G-box sequence CACGTG to CACCCG (Gboxm) abolished the binding of a subset of group I complexes to the −957/−847 probe (Figure 4A, arrow). This mutation was therefore tested in vivo for its effect on *PLHY:luc* expression. As this mutation did not abolish binding of all complexes, we also tested the effects of changes in the 2-bp sequences either upstream or downstream of the core hexamer (class I mutations) or both (class II mutations). Such changes were previously shown to alter the pattern of DNA binding protein complexes forming on G-box containing probes (Williams et al., 1992), and we reasoned that different subsets of G-box binding proteins would be differentially affected by these mutations.

Newly transformed lines carrying the −957 *G-boxm PLHY:luc* construct exhibited rhythmic luminescence but lost expression over time and therefore were not characterized further. Mutations of flanking nucleotides either 5′ or 3′ of the core sequence (class I mutations) or both (class II mutations) caused a twofold reduction in expression levels (Figure 5B). The amplitude of oscillations was reduced under both LD cycles and constant light (Figures 5C and 5D, Table 1). Upon transfer to constant light, a subtle but reproducible broadening of the peak was observed relative to the wild-type construct, with transcription being
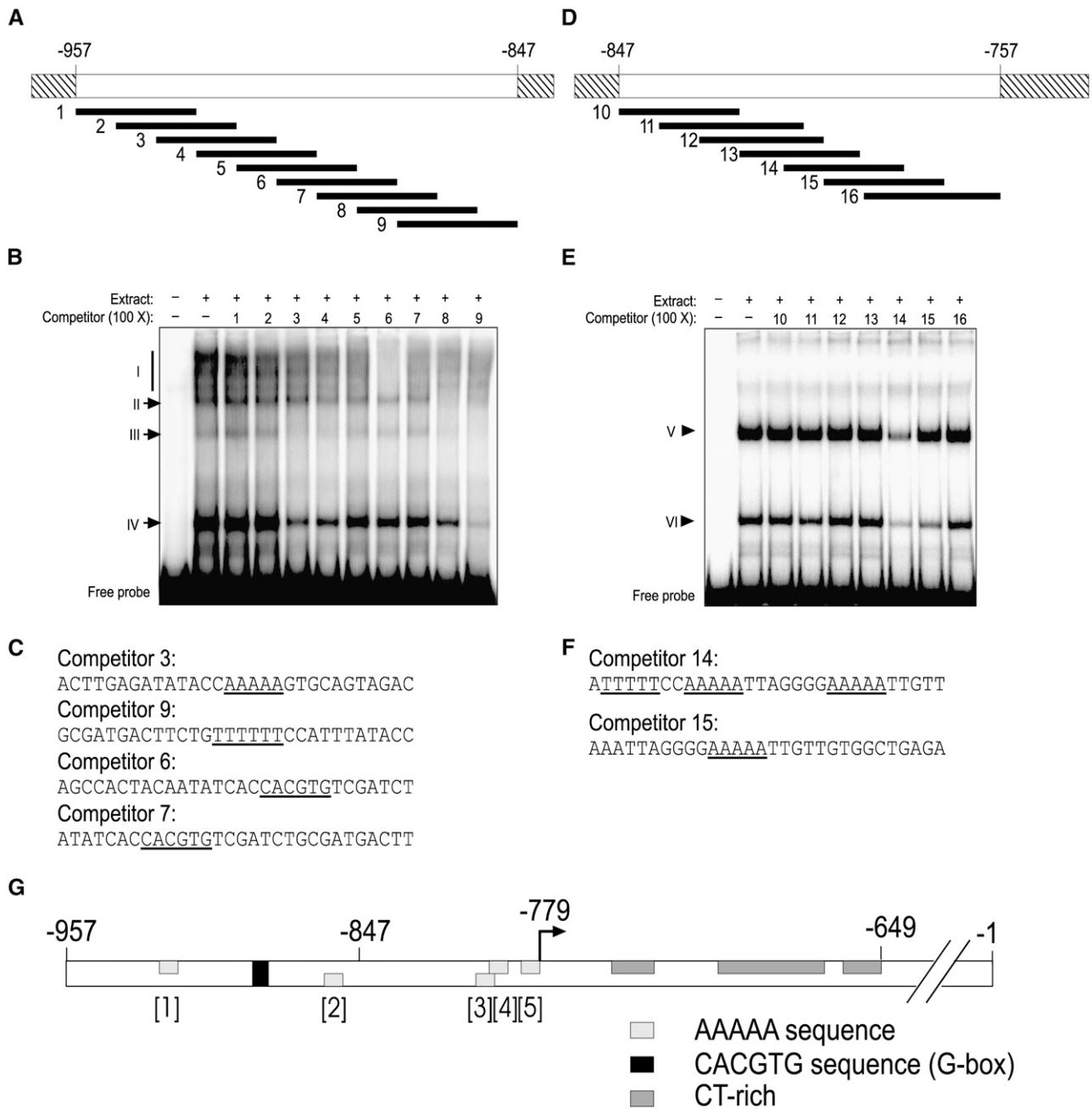
Figure 3. Mapping of Binding Sites for Protein Complexes in the *LHY* Promoter.

(A) and **(D)** Diagrams of EMSA probes (open rectangles). Hatched boxes indicate the incorporation of *Ngo*MIV restriction sites for radiolabeling purposes. Horizontal bars below show the relative positions of the 30-bp oligonucleotides used in competition assays.

(B) and **(E)** EMSAs using the −957/−847 and −847/−757 fragments of the *LHY* promoter as probes. Plant extracts were prepared from tissue harvested at subjective dawn (ZT 24). Different groups of DNA-protein complexes are numbered and indicated by arrows or vertical bars on the left. The + and − symbols at the top of each of the lanes indicate the presence or absence of competitor DNA, and the numbers correspond to specific oligonucleotides used as competitors.

(C) and **(F)** Sequences of oligonucleotides shown to compete for formation of DNA-protein complexes in **(B)** and **(E)**, respectively.

(G) Schematic representation of the LHY promoter showing the relative positions of the G-box (CACGTG), 5A sequences (AAAAA), and CT-rich region. The arrow indicates the position of the transcriptional start site at position −779.
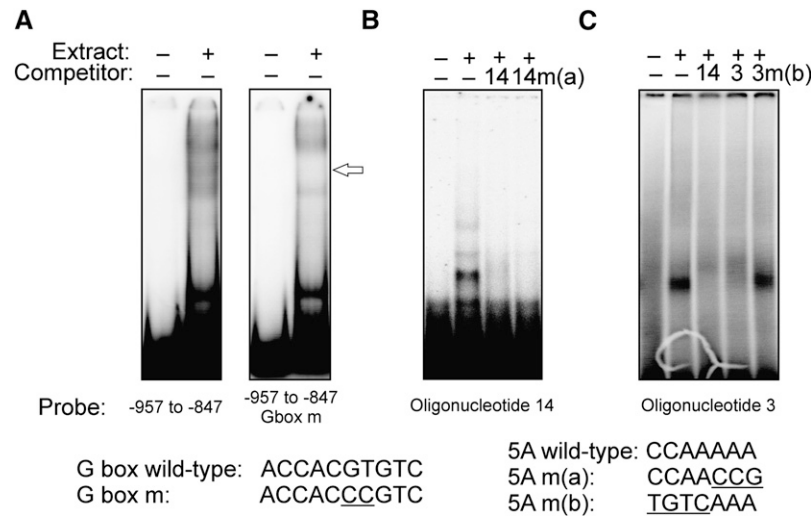
Figure 4. In Vitro Effects of Promoter Mutations.

(A) EMSAs using −957/−847 probes. In the right-hand panel, the G-box sequence (CACGTG) was mutated to CACCCG. The arrow highlights a DNA-protein complex whose binding was abolished by the mutation. Binding of other complexes was significantly reduced as well.

(B) Mutation of CCAAAAA sequences to CCAACCG [m(a)] in the competitor oligonucleotide #14 (Figure 3F) failed to abolish competition for binding to the wild-type probe (oligonucleotide 14).

(C) Mutation of the CCAAAAA sequence to TGTCAAA [m(b)] in the competitor oligonucleotide #3 (Figure 3C) abolished competition for binding to the wild-type probe (oligonucleotide 3).

switched on earlier and returning to trough levels later. Waveform analysis returned lower skewness values than for the wild-type construct, indicative that peaks were less asymmetric (Table 1).

These results demonstrate that the G-box motif contributes to the rhythmic expression pattern of *LHY* both under driven and free-running conditions and suggest a dual role for the G-box motif: first to repress *LHY* expression before dawn and at dusk, restricting expression to a narrow range of phases; and second, to promote *LHY* transcription resulting in high amplitude oscillations.

In Vivo Effects of 5A Mutations

Mutation of all three of the 5A (AAAAA) motifs from oligonucleotide 14 to AACCG failed to alter its ability to compete for binding to the wild-type probe (Competitor 14ma, Figure 4B). Thus, the final three adenosine residues of the 5A motif were not critical for complex formation. Closer examination of competitors 3, 9, 14, and 15 revealed that the AAAAA motif was always preceded by the sequences GG or CC, so another mutation was designed that altered these residues. The change from CCAAAAA to TGTCAAA successfully abolished the ability of competitor 3 to compete for binding to a wild-type probe (Figure 4C). This mutation was therefore introduced into the two 5A motifs flanking the G-box (1,2m constructs) and into all three instances of the motif downstream of position −847 (3,4,5m constructs).

In the context of the −957 PLHY:luc construct, mutation of either 5A sites 1 and 2 or 5A sites 3, 4, and 5 caused a twofold to threefold reduction in expression levels (Figure 6A). Mutation of sites 3, 4, and 5 caused a loss of amplitude in both LD and LL

(Figures 6C and 6E, Table 1). A broadening of the peak in LL was observed, similar to the effect of G-box mutations. Mutation of sites 1 and 2 had a weaker effect. In the context of the −847 PLHY:luc construct, little or no effect on either amplitude or waveform of luminescence was observed when all three 5A sites were disrupted (Figures 6D and 6F). Effects on expression levels were also much less pronounced (Figure 6B). These results suggest that 5A sites contribute to the rhythmic expression pattern of *LHY* but require an element located within the −957/−847 region for this effect. This additional element may be the G-box, since results from Figure 3 suggested a possible interaction between G-box and 5A binding complexes. As suggested above for the G-box, the 5A motif may also have a dual function, mediating both activation and repression of transcription, since it contributed to high expression levels while restricting the timing of expression. Such dual functions of transcription factor binding sites may explain why we failed to detect any rhythmic changes in DNA binding complexes by EMSA. The G-box and 5A sites may be occupied throughout the day and the switch from activating to repressive mode could be achieved through competition between rhythmically expressed activators and repressors binding the same site. Alternatively, these sites may be occupied by constitutive transcription factors that interact with rhythmically expressed coactivators and corepressors.

Contribution of G-Box and 5A Sites to Genome-Wide Regulation of Rhythmic Gene Expression

The effects of mutations in the G-box and 5A motifs were subtle. None of the mutations abolished rhythmic expression from *LHY*
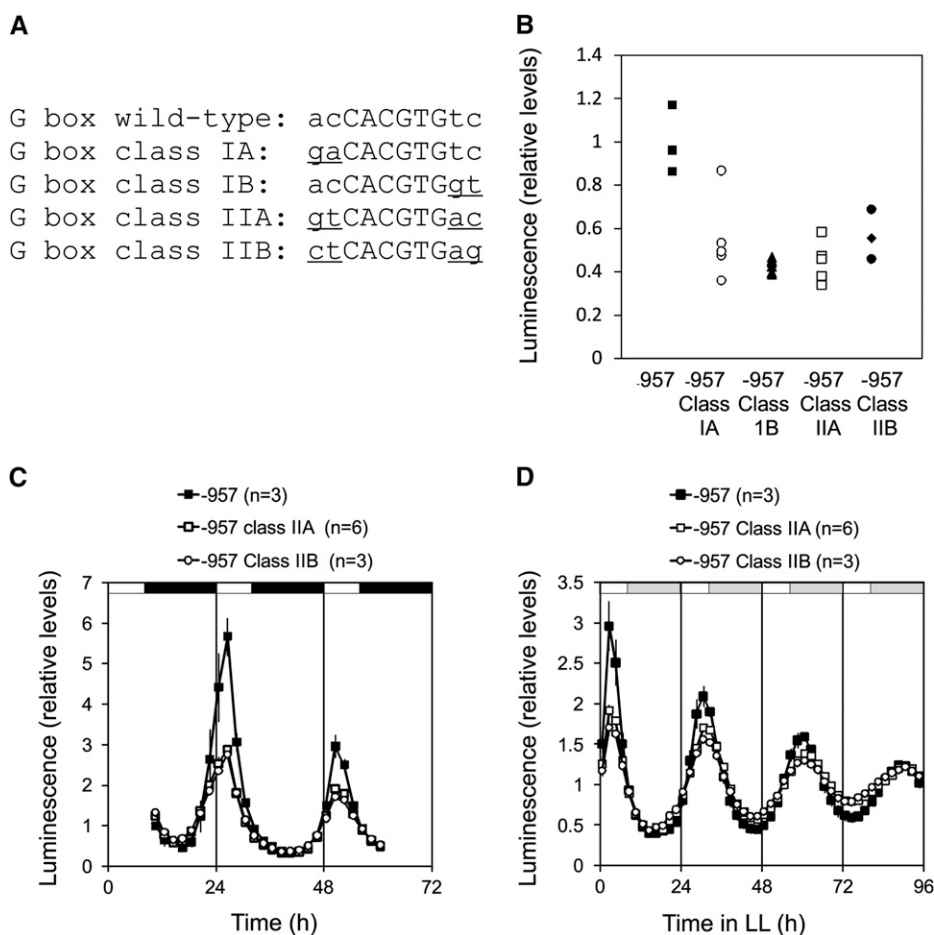
A

```
G box wild-type:  acCACGTGtc
G box class IA:   gaCACGTGtc
G box class IB:   acCACGTGgt
G box class IIA:  gtCACGTGac
G box class IIB:  ctCACGTGag
```

B

C

D

Figure 5. Effects of G-Box Mutations on Expression of *PLHY:luc* Reporter Constructs in Transgenic Plants.

(A) Mutations tested. Nucleotide changes are underlined.
(B) Effects of the mutations on luciferase expression levels. Plants were grown under 8L16D of white light for 7 d and then exposed to 8L16D of red light for 48 h before transfer to constant red light. Each of the data points represents luminescence levels for one transgenic line, averaged over 120 h in constant light and then normalized to average levels for plants carrying the wild-type construct.
(C) Luminescence rhythms under 8L16D cycles of red light.
(D) Luminescence rhythms in constant light.

upstream regions, presumably due to the multiplicity of rhythmic signals feeding into the regulation of this promoter. We therefore sought further evidence for the role of the G-box and 5A motifs in the control of circadian gene expression by testing whether these motifs, either alone or in combination, were enriched within sets of rhythmic genes expressed at specific phases.

Sets of genes associated with different phases under various diurnal or free-running conditions were retrieved from the DIURNAL database (Mockler et al., 2007; Michael et al., 2008). We initially tested for phase-specific enrichment of the hexameric G-box sequence (CACGTG) in these sets of genes, as compared with the full set of 25,516 *Arabidopsis* promoter sequences retrieved from AtcisDB (Molina and Grotewold, 2005). For data sets obtained under entraining LD cycles (16L8D), significant P values were obtained for phases ranging from late night (21 h after dawn, corresponding to zeitgeber time [ZT] 21) to late afternoon (ZT 12) (Figure 7A). Similarly, for data sets obtained

under free-running conditions (constant light or LL), significant P values were obtained from circadian times (CT) 21 to 11 (i.e., from 3 h before subjective dawn to 11 h after subjective dawn) (Figure 7B). The complete analysis and full technical details are shown as part of the Supplemental Methods online.

These results were in good agreement with previous findings showing an association between the presence of a G-box and circadian expression peaking in the daytime (Michael et al., 2008). However, the broad range of expression phases associated with G-box–containing promoters contrasted with the much narrower range of phases associated with evening or morning elements (Michael et al., 2008). This led us to question whether multiple G-box binding factors may be involved that may be active at different phases and have slightly distinct binding specificities. As G-box flanking sequences have been suggested to influence the sign of light responses (Hudson and Quail, 2003), we hypothesized that different sets of G-box flanking nucleotides
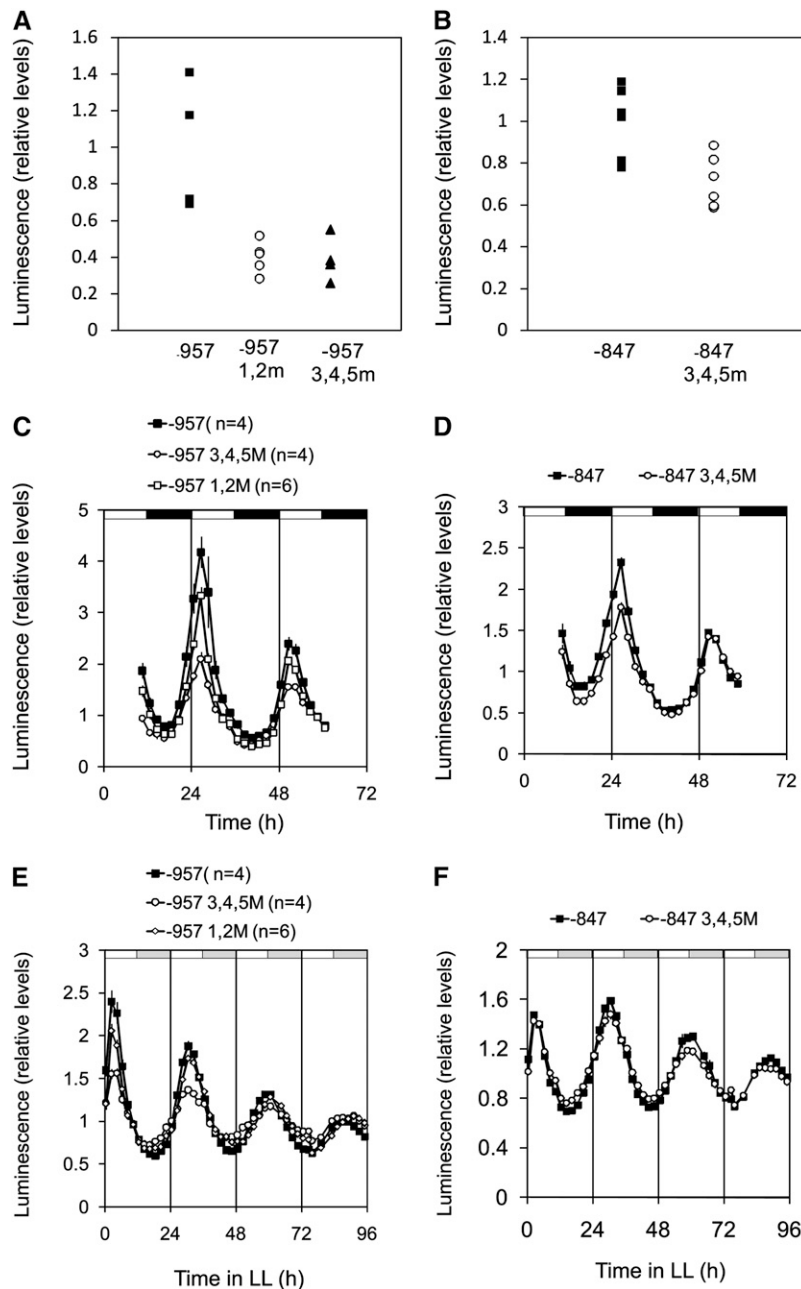
Figure 6. Effects of 5A Mutations on Expression of *PLHY:luc* Reporter Constructs in Transgenic Plants.

(A) and **(B)** Effects of the mutations on luciferase expression levels, in the context of the *−957 PLHY:luc* reporter or of the *−847 PLHY:luc* reporter. The *−957 1,2m* construct carried mutations in flanking the G-box. The *−957* and *−847 3,4,5m* constructs carried mutations in all three 5A sequences located downstream of position −847. Each of the data points represents luminescence levels for one transgenic line, averaged over 120 h in constant light and then normalized to average levels for plants carrying the wild-type construct.
(C) and **(D)** Luminescence rhythms under 12L12D cycles of red light.
(E) and **(F)** Luminescence rhythms in constant red light.

might be associated with expression at different times of the day. Thus, inclusion of 5′ and 3′ flanking bases from the *LHY* promoter (*LHY* G-box; acCACGTGtc) in the analysis returned a narrower range of phases (between ZT 23 and ZT 06 under 16L8D cycles [Figure 7A] and between CT 21 and CT 03 in LL [Figure 7B]). A

different 5′ upstream sequence (gcCACGTG; O G-box) was associated with expression later in the day (Figure 7B).

As a more stringent test for the contribution of G-box flanking sequences to phase specificity, we tested whether among all rhythmic promoters containing the core G-box hexamer, those
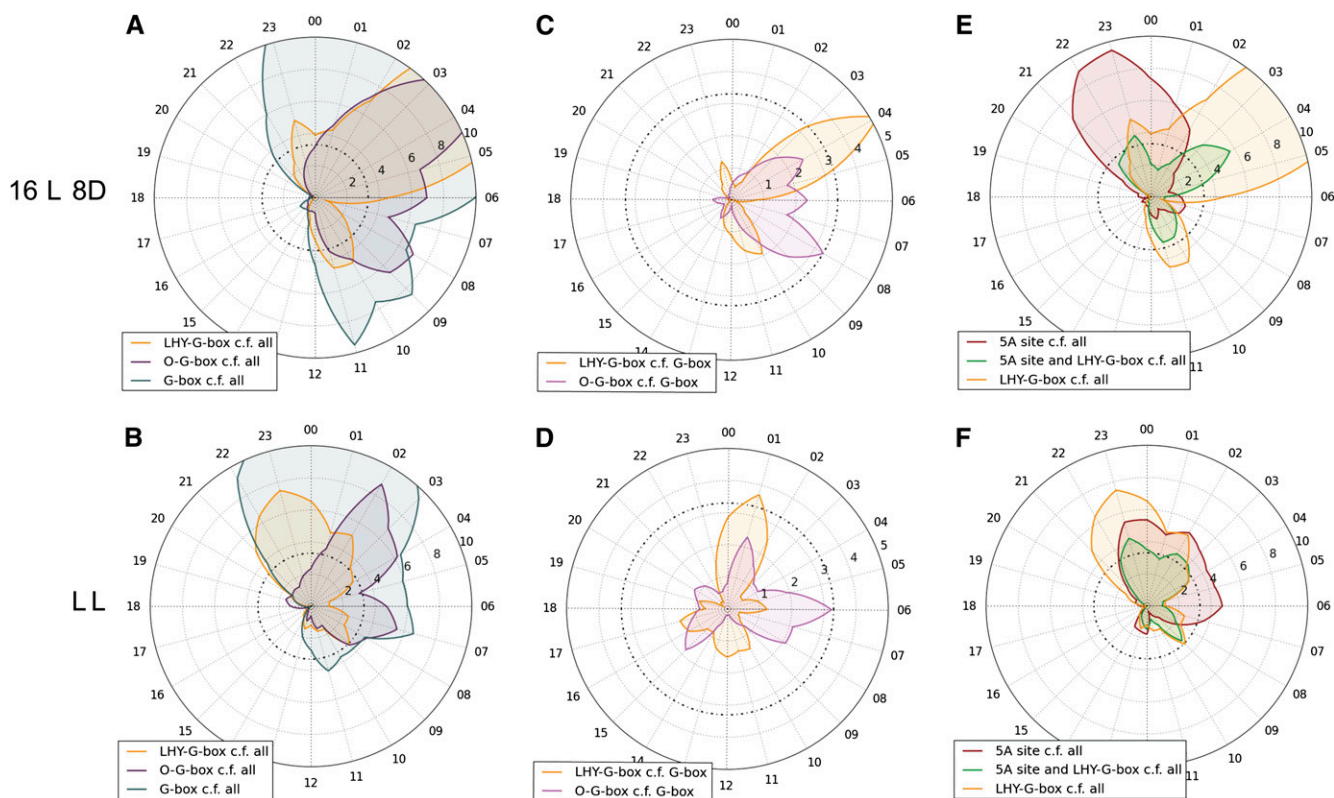
Figure 7. The G-Box and 5A Motifs Are Enriched within Phase-Specific Sets of Promoters.

(A) and **(B)** Phase-specific enrichment for G-box sequences under 16L8D or LL. Enrichment for G-box sequences was tested in sets of genes associated with different phases compared with the full set of 25,516 *Arabidopsis* promoter sequences retrieved from AtcisDB. Enrichment for the G-box hexamer (CACGTG) is shown in dark green, for the LHY G-box (ACCACGTGTC) in yellow, and for the O G-box (GCCACGTG) in purple. Dawn corresponds to time 0. Significance thresholds are indicated by dotted lines.

(C) and **(D)** To test if bases flanking the core G-box hexamer confer phase specificity, we analyzed the enrichment for the LHY G-box and the O G-box against a background of those promoters that contained only the core hexamer. Enrichment for the LHY G-box is shown in yellow and for the O G-box in purple.

(E) and **(F)** Phase-specific enrichment for the 5A motif. Enrichment for the 5A PSSM (in purple) was tested in sets of genes associated with different phases compared with the full set of 25,516 *Arabidopsis* promoter sequences retrieved from AtcisDB. It is compared with the enrichment pattern for the LHY G-box alone, in yellow, and for both motifs, in green.

containing the *LHY* G-box and O G-box were enriched at specific phases (Figures 7C and 7D). Significant P values were obtained for the *LHY* G-box at ZT04 and CT01 under 16L8D and LL, respectively. Similarly, for the O G-box, significant P values were obtained at ZT 08 under 8L16D and at ZT 06 in LL. Analysis under a wider range of environmental cycles (see Supplemental Figure 11 online) suggested that the phase relationship between these two promoter elements varied with photoperiod and in response to temperature cycles, which may indicate control by distinct oscillators. This analysis provided strong and novel evidence that sequences immediately flanking the G-box influence circadian phase specificity.

A new PSSM was generated for the 5A binding site, based on in vitro binding data with an array of wild-type and mutated oligonucleotides (Figure 8). This matrix produced the consensus sequence (A/T)5-CC-(AT)5(T/G)(A/T), a motif related to the CArG box (CC(A/T)$_6$GG bound by the MADS box family of transcription factors (Shore and Sharrocks, 1995). Matches to this PSSM were significantly enriched within rhythmic promoters that were active shortly before dawn under LD conditions (Figure 7E; see Supplemental Figures 12B and 12C online). However, the 5A motif was associated with morning or early afternoon expression under temperature cycles (see Supplemental Figures 12D to 12F online).

To determine how rhythmic signals mediated by the *LHY* G-box and the 5A motif might be integrated at the level of transcriptional activity, we compared the phase enrichment patterns of promoters containing either of these motifs or both. Whether under entraining LD cycles (Figure 7E; see Supplemental Figures 12A to 12C online) or constant light (Figure 7F; see Supplemental Figures 12G and 12H online), the presence of both motifs was associated with phases very similar to the G-box motif alone. However, when plants were exposed to diurnal temperature cycles (LLHC, LDHC, and LL_LLHC), the timing of gene
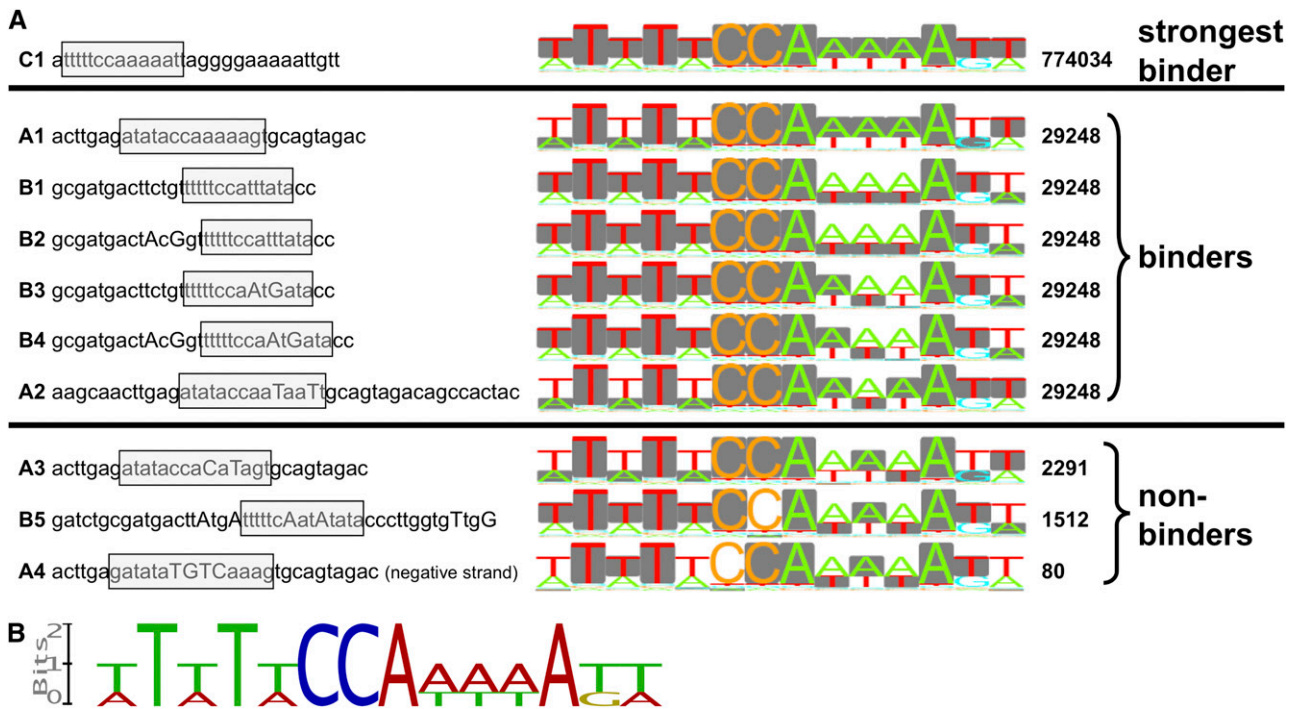
Figure 8. Determination of a New PSSM for the 5A Motif.

A number of oligonucleotides were tested, as shown in Figures 4B and 4C, for their ability to compete for binding of protein complexes to 5A sites in the *LHY* promoter. The PSSM shown in **(B)** was determined by aligning sequences that bound to the same biochemical activity in vitro. **(A)** shows the sequence of the oligonucleotides tested. Sequences A1, B1, and C1 correspond to competitors 3, 9, and 14, respectively. Sequences numbered 2 to 5 correspond to mutated versions of these oligonucleotides. Mutations are indicated by capital letters. Matches to the PSSM are boxed in gray in the corresponding sequence logos, and motif scores are indicated to the right. Scores are the expected number of random k-mers one needs to test in order to find one k-mer that is as close to the PSSM as the site under consideration. The strongest binder in vitro perfectly matches the weight matrix and scores highest. All other binders have good matches with weight matrix and a high score. Nonbinders have significant mismatches and distinctly low scores.

[See online article for color version of this figure.]

expression seemed determined primarily by the presence of the 5A motif (see Supplemental Figures 12D to 12F online). Further experimentation will be required to understand how these regulatory sequences interact to modulate the timing of *LHY* transcription.

The G-Box and 5A Sites Are Conserved within the Promoters of Orthologous Genes

Promoter regions that are functionally important are expected to be conserved during evolution. We applied a recently developed comparative genomics technique (E. Picot, I.A. Carre, and S. Ott, unpublished data) to identify regulatory modules that are conserved between *LHY* orthologs from distant species. Figure 9 shows that the region comprising the G-box and 5A motifs exhibits a significant level of conservation between *Arabidopsis*, grapevine (*Vitis vinifera*), castor bean (*Ricinus communis*), and poplar (*Populus trichocarpa*). Strikingly, this region of significant conservation matches the functional region we defined by purely experimental means. A G-box motif was present at a conserved position in all four species. Interestingly, the fourth base of the core hexamer (CACGTG) was not conserved, indicating a very loose binding requirement at this position for the cognate tran-

scription factor. Matches to the 5A PSSM were identified in all four promoters, but their multiplicity and positions were not always conserved. This is not surprising as evolutionary pressures would be expected to be reduced for motifs that are present in multiple, partially redundant copies. Loss and/or relocation of some sites may not abolish regulation.

The alignment shown in Figure 9B highlights three other regions of remarkable conservation. These may correspond to transcription factor binding sites that were not detected by our biochemical analysis, possibly because of masking by G-box and 5A binding complexes. In *Arabidopsis*, grapevine, and castor bean, conserved regions 1 and 3 contained inverted copies of the sequence CAGCCAC, and the perfect duplication of this sequence provides further evidence for its functional importance in the regulation of *LHY* transcription. A CT-rich region was also present in all four orthologous promoters, although alignments were poor and consequently not shown.

The Transcription Factor FLC Binds to the *LHY* Promoter

The MADS box transcription factor *FLC* plays a major role in the regulation of flowering time in *Arabidopsis* but has also been
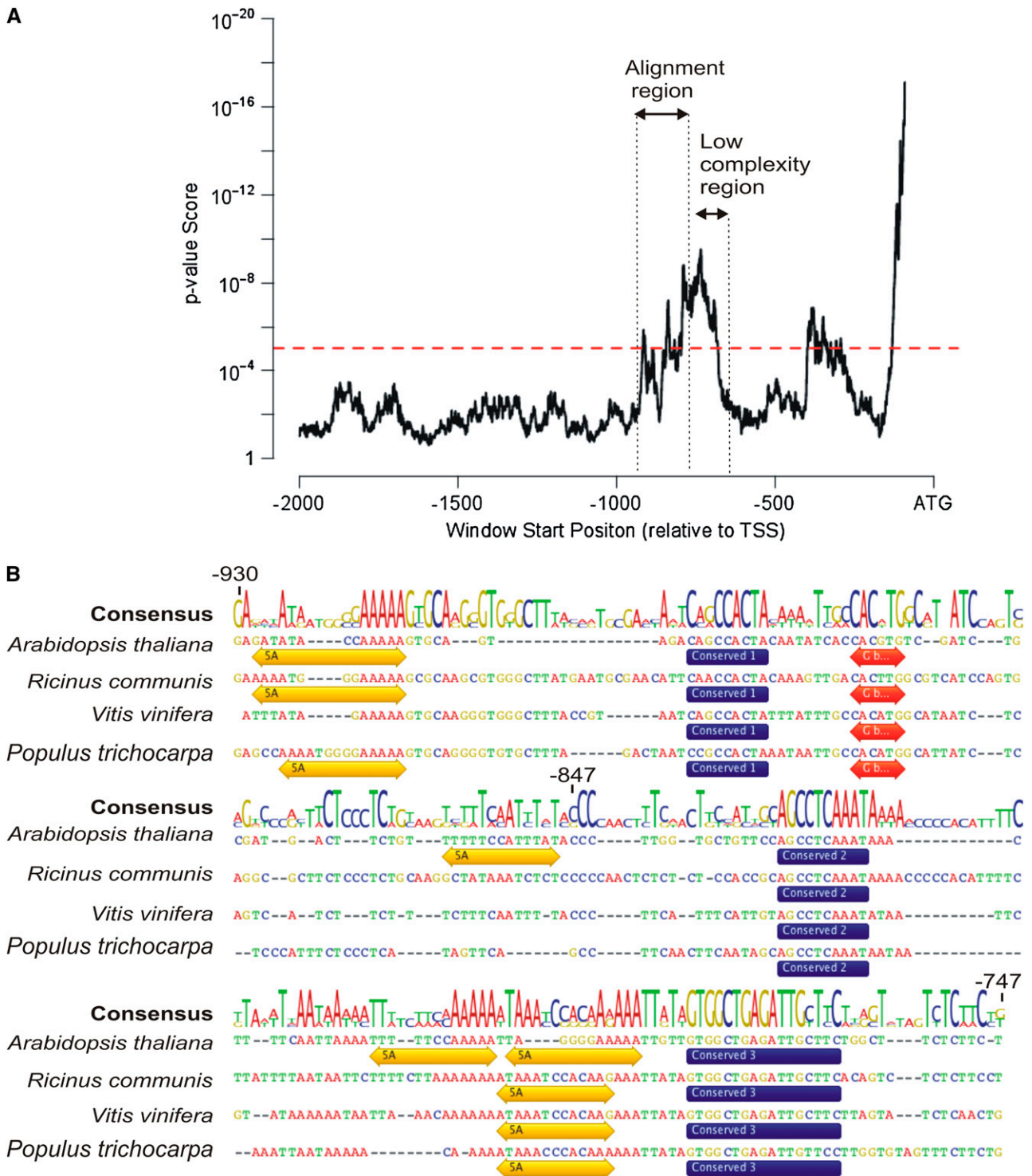
Figure 9. Identification of Evolutionarily Conserved Sequences within the Promoter of *LHY*.

(A) *LHY* conservation profile. Cumulative conservation profile between the *Arabidopsis LHY* promoter and orthologous promoters from grapevine, castor bean, and poplar. Two thousand bases upstream of the translational start site of the *LHY* gene were aligned using the ReMo algorithm with a 90-base window length and a 1-base step width. The dotted red line indicates the significance threshold of P = 0.00001. Peaks above this threshold indicate that the window has a highly conserved match in the other species. Sequences between positions −930 and −747 aligned well and are shown below in **(B)**. Sequences between positions −747 and −679 consisted mostly of CTT repeats. Due to their low complexity, they gave a high conservation score but did not align well.

shown to modulate the period of the circadian clock and contribute to its temperature compensation (Swarup et al., 1999; Edwards et al., 2006). As the FLC binding site in the *SOC1* promoter (5′-TTTTCCAAAATAAGTAAA-3′) contains a perfect match to our PSSM for the 5A motif (Helliwell et al., 2006), we tested whether the FLC protein might interact with the *LHY* promoter. Chromatin immunoprecipitation (ChIP) experiments were performed using an antibody to the FLC protein. Precipitation of chromatin from *35S:FLC-FLAG* transgenic lines (Helliwell et al., 2006) resulted in a fourfold to fivefold enrichment of *LHY* promoter sequences over nontarget, control DNA (Figure 10A). A weak enrichment (up to twofold) was obtained with chromatin samples from wild-type (C24) plants, whereas no enrichment was observed with chromatin from plants lacking FLC (*flc20* mutant).

These results indicated that the FLC protein can bind to the *LHY* promoter in planta. For comparison, we tested binding of FLC to its known target sequence in the *SOC1* promoter (Figure 10C). *SOC1* promoter sequences were enriched >50-fold in C24 ChIP samples, showing that the interaction of FLC with *SOC1* is more frequent than with *LHY* by at least one order of magnitude.

DISCUSSION

Here, we combined traditional analysis of deletion constructs, biochemical analysis, and bioinformatics to identify regulatory elements within the *LHY* promoter and provide evidence for their function in mediating rhythmic gene expression. This work highlights the vast potential of comparative genomic approaches to identify functional elements within promoters but also demonstrates the value of more traditional methods for identification of degenerate transcription factor binding sites.

A G-Box Motif Contributes to the Rhythmic Expression of *LHY*

We showed that the G-box motif (CACGTG) contributes to the level and temporal pattern of *PLHY:luc* transcription in constant light. Mutations that altered two nucleotides either upstream, downstream, or both upstream and downstream of the core hexamer sequence all affected the amplitude and timing of the *LHY* expression rhythm. These results suggested that the G-box sequence might mediate a rhythmic signal to the *LHY* promoter and highlighted the functional importance of immediate flanking sequences. A statistical analysis of time of day–specific enrichment within rhythmic promoters provided further evidence for this. The G-box hexamer was shown to be associated with daytime expression at a broad range of phases. By contrast, a 10-bp sequence from the *LHY* promoter comprising the G-box
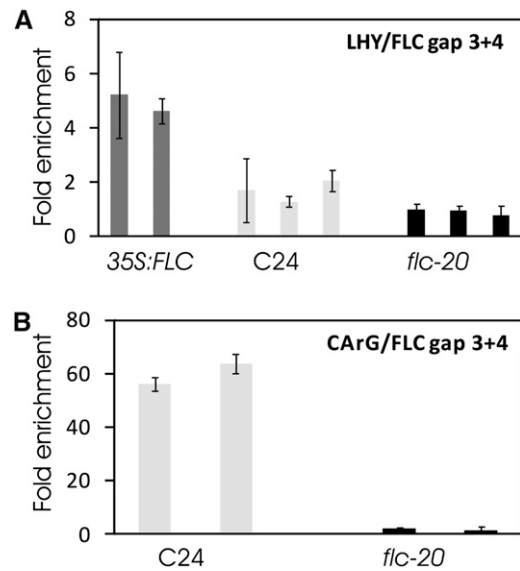

Figure 10. ChIP Experiments Using an Antibody to the FLC Protein.

Each bar represents the mean of two ChIP experiments from one biological sample. Error bars indicate SD. Replicates from two or three independent experiments are shown as adjacent bars. Fold enrichments for various sequences were calculated relative to a region in the *FLC* gene that is not enriched by ChIP with FLC antibodies.
(A) Enrichment for *LHY* promoter DNA in *35S:FLC-FLAG*, wild-type (C24), and *flc20* mutant lines.
(B) Enrichment for *SOC1* promoter DNA in *35S:FLC-FLAG* and *flc20* mutant samples.

and immediate upstream and downstream flanking sequence was associated with morning-specific expression. A G-box with distinct flanking sequences was associated with expression at different circadian phases. This demonstrates that contextual information is required to properly evaluate the contribution of the G-box motif to circadian-regulated gene expression. Thus, the CCAC sequence found in a dawn-specific G-box, gCCACgtg, was suggested to form the core of a morning-specific module shared with the morning element, nCCCACacn (Michael et al., 2008). The importance of G-box flanking sequences was also suggested with regard to light responses, as the AcCACGTGtca sequence was found to be enriched within sets of Phytochrome A–induced genes, whereas a distinct 3′ flanking sequence (cCACGTGggag) was associated with light repression (Hudson and Quail, 2003).

The G-box motif is recognized by a family of basic helix-loop-helix transcription factors, including PHYTOCHROME INTERACTING FACTOR3 (PIF3) and the related PIF3-LIKE (PIL)

Figure 9. (continued).

(B) Alignment of conserved regions of the *Arabidopsis*, grapevine, and castor bean promoters. Red arrows highlight a conserved G-box motif. Yellow arrows indicate 5A sites, identified according to the PSSM in Figure 5 allowing two mismatches in the least conserved bases and no mismatches in the highly conserved bases. Blue bars indicate highly conserved regions that don't correspond to any known PSSM in the TRANSFAC database and may correspond to novel transcription factor binding sites.
[See online article for color version of this figure.]

proteins (Leivar et al., 2008). PIF3 and several PILs were shown to interact with the circadian clock component, TOC1, in yeast two-hybrid assays (Yamashino et al., 2003). This suggested that PIF3 or PILs might mediate the rhythmic effect of TOC1 on *LHY* transcription and therefore play a role as a component of the clock mechanism. However, plants either lacking or overexpressing PIF3 showed no alterations in their circadian rhythms so this remains to be explored (Oda et al., 2004; Viczián et al., 2005).

Is There a Role for a MADS Box Transcription Factor in the Regulation of LHY Transcription?

In addition to the G-box, we identified a more degenerate regulatory sequence that we named the 5A motif. This sequence was significantly enriched within sets of rhythmically expressed genes and in a phase-specific manner, suggesting that it does contribute to the timing of circadian-regulated gene expression. Our comparative genomic analysis showed the presence of similar sequences in the promoters of *LHY* orthologs from grapevine, poplar, and castor bean plants. A new PSSM generated for this motif suggested that it might be related to CArG boxes, which are bound by transcription factors of the MADS box family.

In support of this hypothesis, we showed that the MADS box transcription factor FLC can bind to the *LHY* promoter in vivo. This interaction was much weaker than with the *SOC1* promoter and was only detected reliably in FLC-overexpressing plants. It is not clear at this point whether the weak interaction detected in wild-type plants has any functional significance. The interaction detected in *35S:FLC-FLAG* plants may only reflect ectopic binding of FLC to sites that are normally bound by other MADS box transcription factors (Helliwell et al., 2006). Extensive low affinity transcription factor interactions have been observed in yeast, and many predicted low-affinity binding sites were conserved during evolution, suggesting that they are functionally important (Tanay, 2006). It is therefore plausible that low-affinity binding of FLC to the *LHY* promoter might modulate its activity. Alternatively, weak enrichment for *LHY* sequences in ChIP experiments might reflect high-affinity binding of FLC that is limited to a few cell types or to specific times of the day. Expression of FLC is arrhythmic and fairly ubiquitous, but since MADS box transcription factors bind as dimmers, it could be that FLC requires a partner that is expressed in a cell-specific or circadian-regulated fashion. Further work will be required to test whether FLC binding to the *LHY* promoter exhibits such regulation. FLC contributes to the temperature compensation of the circadian clock (Edwards et al., 2006), and changes in *LHY* expression levels have also been implicated in this process (Gould et al., 2006). It is tempting to speculate that temperature-dependent changes in *LHY* expression levels might be mediated through FLC binding to 5A sites within the *LHY* promoter.

Additional Promoter Elements Identified by a Comparative Genomic Analysis

Neither the 5A or G box mutations within the −957 to −847 region reproduced the same effect as the deletion of the entire region on the phase and waveform of transcription. This may

reflect the action of additional regulatory motifs within this region. Furthermore, the disruption of 5A sites in the context of the −847/1 promoter only had a subtle effect, suggesting that additional sequences mediate circadian regulation of this minimal promoter fragment.

The comparative analysis of orthologous *LHY* promoters identified a conserved sequence (conserved region 1; CAGC-CACTA) within the −957/−847 region that may account for its effect on phase. Two further blocks of conserved sequence located downstream of position −847 (conserved regions 2 and 3; AGCCTCAAATA and GTGGCTGAGATTGCTTC) may also play a role in the rhythmic expression of the −847 pLHY: luc construct in the absence of functional 5A sites. None of these sequences contains any known rhythmic elements, although conserved regions 1 and 3 both comprise the SORLIP 1 sequence (GCCAC) and contain incomplete matches to the morning element (CCAC or GTGG in reverse orientation). Their role in the control of *LHY* transcription remains to be characterized. Sequences from positions −747 to −679 are largely composed of trinucleotide (CTT) repeats. This region is predicted to comprise 16 binding sites for GA binding, basic pentacysteine proteins. Cooperative binding of these proteins to multiple sites has been shown to induce conformational changes in promoter regions and was hypothesized to allow recruitment of additional regulatory complexes (Pruneda-Paz et al., 2009). The role of the CT-rich region in the regulation of *LHY* transcription is currently being investigated.

Conserved and Divergent Aspects of *LHY* and *CCA1* Regulation

Since *LHY* and *CCA1* exhibit similar patterns of dawn-specific transcription, we expected at least some transcription factor binding sites to be conserved between the two promoters. Indeed, *CCA1* upstream sequences comprised at least 14 matches to the 5A PSSM (see Supplemental Figure 14 online) and two G-boxes. The 3′ flanking sequences of the most distal G-box were distinct from those of the *LHY* G-box, but the proximal G-box was a good match and was consistent with the similar phase of expression of those two genes. Several CT-rich regions were identified in *CCA1*, although these were not as extensive as in *LHY*. Strikingly, the *CCA1* promoter contained nearly perfect matches to the three evolutionarily conserved regions of unknown function identified in the *LHY* promoter, while the rest of the promoter was highly divergent. Therefore, *LHY* and *CCA1* share short stretches of promoter sequence in which mutations are under strong negative selective pressure. However, the *LHY* promoter does not contain a binding site for CHE (GGNCCCAC), a rhythmic transcriptional repressor binding to the *CCA1* promoter. This suggests that CHE may not directly regulate *LHY* transcription. This difference may in turn account for the differential regulation of *LHY* and *CCA1* transcription under extreme high or low temperatures, which is believed to contribute to the temperature compensation of the clock (Gould et al., 2006).

Multioscillator Control of *LHY* Transcription

The elements controlling rhythmic transcription of *LHY* proved to be highly redundant, and none of the mutations tested abolished

circadian rhythmicity. These observations contrast with those for the *TOC1* or *CAT3* promoters, where mutation of a single evening element was sufficient to severely disrupt or completely abolish rhythmic transcription (Alabadi et al., 2001; Michael and McClung, 2002). The greater complexity of *LHY* regulation fits well with its proposed role at the core of the circadian network and at the intersection of multiple oscillatory feedback loops. Moreover, altered patterns of photoperiod-dependent phase adjustments as a result of the −957/−847 deletion suggested that input from one out of several oscillatory feedback loops may have been disrupted. The identification of regulatory elements within the *LHY* promoter and their assignment to different regulatory feedback loops will now provide tools to investigate the functional implications of this multioscillator coupling.

METHODS

Generation of *LHY* Deletion Constructs

The *PLHY:LUC1* construct has been described previously (Kim et al., 2003). The 3′ UTR sequence of LHY was amplified from the cDNA using forward primer 5′-GGGGGCTAGCACATGACAGACTTGGAGGTA-3′ and a reverse primer to the T7 promoter sequence. The PCR product was inserted between the *Eco*RV and *Not*I sites of the *PLHY:LUC1* construct to form *PLHY:LUC3*. The *PLHY:LUC5* construct was created by synthesizing a double-stranded DNA linker (forward strand, 5′-AAAT-ATTCTCTCTCAACCAAAATATTCGATAC-3′; reverse strand, 5′-CATGG-TATCGAATATTTTGGTTGAGAGAGAATATTT-3′) corresponding to the *nos* 5′ UTR, and inserting it between the *Dra*I and *Nco*I sites of *PLHY:LUC1*. The 5′ deletion constructs are described in the supplemental data online as part of the full-length upstream sequence of *LHY* (see Supplemental Figure 1 online). For construction of *35S:LHY5′UTR:luc*, the full-length *35S* promoter was PCR amplified from the *35S:luc* plasmid (Hellens et al., 2000) using forward primer 5′-CCCAAGCTTATCGTAC-CCCTATCC-3′ and reverse primer 5′-GAGCCCGGGCTGTCCTCTC-CAAAT-3′, digested with *Hin*dIII and *Sma*I and ligated into the *Hin*dIII and *Dra*I sites of *PLHY:luc1*. All constructs were sequenced prior to transformation into *Agrobacterium* (c58).

Transgenic Plants

All constructs were transformed into the wild-type plants of the Ws ecotype by the floral dip method (Clough and Bent, 1998). T2 seeds were harvested from at least six T1 plants resulting from independent transformations.

Luciferase Imaging

For circadian experiments, T2 transgenic plants carrying *PLHY:luc* reporter constructs were sown on Murashige and Skoog agar medium containing 3% sucrose, in clusters of 20 to 40 seedlings (one cluster per transgenic line). They were grown under 24-h LD cycles of white light for 7 d at 22°C and then sprayed with 5 mM luciferin as previously described (Millar et al., 1992) and transferred to the imaging chamber. They were maintained under equivalent LD cycles of monochromatic red or blue light (10 μM·m^{-2}·s^{-1}) for another 72 h prior to transfer to constant light or darkness. Luminescence was monitored by digital imaging of plants using either the ORCAII c4742-98 CCD camera system (Hamamatsu) or a liquid nitrogen–cooled TEK 512x512DB CCD with an ST138 controller (Princeton Instruments). All experiments were repeated at least three times with similar results.

Luciferase expression data were further analyzed using the BRASS 3.0 software (Locke et al., 2005) (available from http://millar.bio.ed.ac.uk/Downloads.html) using the FFT-NLLS algorithm (Plautz et al., 1997). Trough times in Figure 2 were determined manually.

EMSAs

Whole-cell plant extracts were prepared according to Carré and Kay (1995). Binding reactions contained 10 to 20 fmol of radiolabeled probe and 5 μg of protein extract in 12 μL total volume of binding buffer (25 mM HEPES-NaOH, pH 7.5, 50 mM KCl, 0.1 mM EDTA, 20% glycerol, 1 mM DTT, and 1% [v/v] plant protease inhibitor cocktail). Incubations were performed at room temperature for 30 min and then DNA–protein complexes were resolved by electrophoresis on a 6% acrylamide: Bis-acrylamide, 0.25× TBE gel at 200 V. Gels were exposed overnight on a phosphor imager screen.

Comparative Genomic Analysis of Orthologous Promoters

We used a comparative genomics technique for detecting short non-coding conserved sequences with high sensitivity (E. Picot, I.A. Carre, and S. Ott, unpublished data). First, for each 90-mer in the *Arabidopsis thaliana* promoter, the best-aligning 90-mer in each of the orthologous promoters was determined. Then for each group of 90-mers, a statistical test is done probing for evidence of orthology. As only distant species are used in which nonfunctional regions have sufficiently diverged, evidence for orthology implies evidence for conservation. Plotting the P values for all *Arabidopsis* 90-mers along their position in the promoter is termed "conservation profile" in the context of this work. Sequence alignments were generated using Geneious v4.0.

ChIP

ChIP analysis for FLC binding was performed essentially as described (Helliwell et al., 2006). Briefly, 1 g of tissue (whole seedlings or shoot tips) was cross-linked in 15 mL, 1% (v/v) formaldehyde in 1× PBS for 1 h. Extracts were prepared by grinding cross-linked material in liquid nitrogen and adding to 5 mL of lysis buffer (50 mM HEPES, pH 7.5, 150 mM NaCl, 1% [v/v] Triton X-100, 0.1% [w/v] deoxycholic acid, 0.1% [w/v] sodium dodecyl sulfate, and 1× plant protease inhibitor mix [Sigma-Aldrich]), sonicating 4 times for 15 s, and centrifuging for 20 min at 20,000*g* to remove debris. For each immunoprecipitation reaction, 1 mL of lysate was incubated with 2 μL of FLC antiserum (Robertson et al., 2008) for 2 h at 4°C; 50 μL of protein A agarose was then added and the sample incubated a further 1 h before the agarose beads were washed and DNA eluted as described (Helliwell et al., 2006). ChIP was performed on C24 wild type, *flc20* mutants (Helliwell et al., 2006), and C24 transformed with *35S:FLC-FLAG* (Helliwell et al., 2006). Plants were grown for 12 d in a 16L:8D photoperiod and harvested ∼4 h after dawn. The immunoprecipitations were replicated two times for apex and three times for seedling samples.

ChIP samples were analyzed by quantitative real-time PCR using primers 5′-CGTGTCGATCTGCGATGACT-3′ and 5′-AAAAGTTTATTT-GAGGCTGGAACAG-3′ for *LHY*, 5′-GGCATTTCCATCCAACTAG-3′ and 5′-ATCAGTCAGTATACACAGC-3′ for the *SOC1* CArG box, and 5′-TCA-GAGCTTTTGACTGATGATCCT-3′ and 5′-CCCTTGTCTTTTACCGCT-TCTTC-3′ as a nonenriched control (these primers amplify a region in the *FLC* gene that is not enriched by ChIP with FLC antibodies). Quantitative PCR was performed using an Applied Biosystems 7900HT instrument. Reaction conditions were 1× Platinum Taq buffer (Invitrogen), 3.5 mM MgCl$_2$, 0.2 mM deoxynucleotide triphosphate, 1:20,000 SYBR Green, and 0.05 units/μL Platinum Taq. Reaction conditions were 95°C for 5 min followed by 40 cycles of 95°C for 15 s, 60°C for 15 s, and 72°C for 30 s. Product sizes were verified by agarose gel electrophoresis and

dissociation curve analysis. Amplicons in ChIP samples were quantified using a standard curve of total genomic DNA. All reactions were run in triplicate and the mean quantity of each amplicon calculated for each ChIP sample. The mean enrichment of *LHY* or *SOC1* sequences compared with the control sequences was calculated across the replicate ChIP samples.

Accession Numbers

Sequence data for the genes described in this article can be found in the Arabidopsis Genome Initiative data library under the following accession numbers: *LHY* (At1g01060), *CCA1* (At2g46830), *FLC* (At5g10140), *SOC1* (At2g45660); in the Gramene database for Pt *LHY* (estExt_Genewise1_v1.C_LG_XIV1950) and Vv *LHY* (GSVIVG00026185001); or in the castor bean genome database for Rc *LHY* (30,076.m004464).

Supplemental Data

The following materials are available in the online version of this article.

   Supplemental Figure 1. 5′ Upstream Sequences of the *LHY* Gene.

   Supplemental Figure 2. Expression Levels of 5′ Deletion Constructs.

   Supplemental Figure 3. Rhythmic Expression of *PLHY:luc* Deletion Constructs.

   Supplemental Figure 4. Comparison of mRNA Expression Patterns for the *-847PLHY:luc* Transgene and for the Endogenous *LHY* Gene.

   Supplemental Figure 5. 5′ and 3′ Untranslated Regions of *LHY* Don't Make Essential Contributions to the Rhythmic Pattern of Gene Expression.

   Supplemental Figure 6. The *-847PLHY:luc* Construct Contains Sequences Mediating Negative Autoregulation of *LHY* Expression.

   Supplemental Figure 7. Quantification of Competition EMSAs Shown in Figure 3.

   Supplemental Figure 8. The Same Biochemical Activity Binds Multiple Sites in the *LHY* Promoter, All Containing the AAAAA Motif.

   Supplemental Figure 9. PSSMs Used in the Phase Enrichment Analysis.

   Supplemental Figure 10. Phase-Specific Enrichment for G-Box Sequences.

   Supplemental Figure 11. Phase-Specific Enrichment for G-Box Flanking Sequences.

   Supplemental Figure 12. Phase-Specific Enrichment for the 5A Motif.

   Supplemental Figure 13. Analysis of Cooperativity between the 5A and the G-Box Motifs.

   Supplemental Figure 14. 5′ Upstream Sequences of *CCA1* Contain Matches to Regulatory Elements Identified in the *LHY* Promoter.

   Supplemental Table 1. Number of Genes Comprising Matches to the PSSMS.

   Supplemental Table 2. Rhythmic Data Sets Used in the Phase Enrichment Analysis.

   Supplemental Table 3. Number of Genes Included in the Analysis of Cooperativity.

   Supplemental Methods.

ACKNOWLEDGMENTS

Author Contributions

J.-Y.K. generated the initial set of deletion constructs and introduced them into plants. M.S. analyzed transgenic lines, identified DNA binding complexes and their cognate binding sites using biochemical methods, introduced mutations in the reporter constructs, and analyzed the mutated constructs in transgenic plants. S.O. and E.P. carried out the comparative genomic analysis of the *LHY* promoter, identified candidate transcription factor binding sites using a bioinformatics approach, and produced the PSSM for the 5A motif. J.R. carried out the analysis of phase-specific enrichment for G-box and 5A motifs. C.H. carried out the ChIP analysis of FLC interaction with the *LHY* promoter. I.A.C. designed and supervised the research and prepared the manuscript.

REFERENCES

Alabadi, D., Oyama, T., Yanovsky, M.J., Harmon, F.G., Mas, P., and Kay, S.A. (2001). Reciprocal regulation between *TOC1* and *LHY/CCA1* within the *Arabidopsis* circadian clock. Science **293:** 880–883.

Alabadi, D., Yanovsky, M.J., Mas, P., Harmer, S.L., and Kay, S.A. (2002). Critical role for *CCA1* and *LHY* in maintaining circadian rhythmicity in *Arabidopsis.* Curr. Biol. **12:** 757–761.

Carré, I.A., and Kay, S.A. (1995). Multiple DNA-protein complexes at a circadian-regulated promoter element. Plant Cell **7:** 2039–2051.

Clough, S.J., and Bent, A.F. (1998). Floral dip: A simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana.* Plant J. **16:** 735–743.

Dodd, A.N., Salathia, N., Hall, A., Kevei, E., Toth, R., Nagy, F., Hibberd, J.M., Millar, A.J., and Webb, A.A.R. (2005). Plant circadian clocks increase photosynthesis, growth, survival, and competitive advantage. Science **309:** 630–633.

Edwards, K., Anderson, P., Hall, A., Salathia, N., Locke, J., Lynn, J., Straume, M., Smith, J., and Millar, A. (2006). *FLOWERING LOCUS C* mediates natural variation in the high-temperature response of the *Arabidopsis* circadian clock. Plant Cell **18:** 639–650.

Gould, P.D., Locke, J.C.W., Larue, C., Southern, M.M., Davis, S.J., Hanano, S., Moyle, R., Milich, R., Putterill, J., Millar, A.J., and Hall, A. (2006). The molecular basis of temperature compensation in the *Arabidopsis* circadian clock. Plant Cell **18:** 1177–1187.

Guo, A., He, K., Liu, D., Bai, S., Gu, X., Wei, L., and Luo, J. (2005). DATF: A database of *Arabidopsis* transcription factors. Bioinformatics **21:** 2568–2569.

Harmer, S.L. (2009). The circadian system in higher plants. Annu. Rev. Plant Biol. **60:** 357–377.

Harmer, S.L., Hogenesch, J.B., Straume, M., Chang, H.S., Han, B., Zhu, T., Wang, X., Kreps, J.A., and Kay, S.A. (2000). Orchestrated transcription of key pathways in *Arabidopsis* by the circadian clock. Science **290:** 2110–2113.

Harmer, S.L., and Kay, S.A. (2005). Positive and negative factors confer

phase-specific circadian regulation of transcription in *Arabidopsis.* Plant Cell **17:** 1926–1940.

Hellens, R.P., Edwards, E.A., Leyland, N.R., Bean, S., and Mullineaux, P.M. (2000). *pGreen*: A versatile and flexible binary Ti vector for *Agrobacterium*-mediated plant transformation. Plant Mol. Biol. **42:** 819–832.

Helliwell, C.A., Wood, C.C., Robertson, M., James Peacock, W., and Dennis, E.S. (2006). The *Arabidopsis* FLC protein interacts directly *in vivo* with *SOC1* and *FT* chromatin and is part of a high-molecular-weight protein complex. Plant J. **46:** 183–192.

Hudson, M.E., and Quail, P.H. (2003). Identification of promoter motifs involved in the network of phytochrome A-regulated gene expression by combined analysis of genomic sequence and microarray data. Plant Physiol. **133:** 1605–1616.

Kim, J.Y., Song, H.R., Taylor, B.L., and Carre, I.A. (2003). Light-regulated translation mediates gated induction of the *Arabidopsis* clock protein LHY. EMBO J. **22:** 935–944.

Leivar, P., Monte, E., Al-Sady, B., Carle, C., Storer, A., Alonso, J.M., Ecker, J.R., and Quail, P.H. (2008). The *Arabidopsis* phytochrome-interacting factor PIF7, together with PIF3 and PIF4, regulates responses to prolonged red light by modulating phyB levels. Plant Cell **20:** 337–352.

Locke, J., Southern, M., Kozma-Bognár, L., Hibberd, V., Brown, P., Turner, M., and Millar, A. (2005). Extension of a genetic network model by iterative experimentation and mathematical analysis. Mol. Syst. Biol. **1:** 13.

Locke, J.C., Kozma-Bognár, L., Gould, P.D., Fehér, B., Kevei, E., Nagy, F., Turner, M.S., Hall, A., and Millar, A.J. (2006). Experimental validation of a predicted feedback loop in the multi-oscillator clock of *Arabidopsis thaliana.* Mol. Syst. Biol. **2:** 59.

Martinez-Garcia, J.F., Huq, E., and Quail, P.H. (2000). Direct targeting of light signals to a promoter element-bound transcription factor. Science **288:** 859–863.

Mas, P., Alabadi, D., Yanovsky, M.J., Oyama, T., and Kay, S.A. (2003). Dual role of *TOC1* in the control of circadian and photomorphogenic responses in *Arabidopsis.* Plant Cell **15:** 223–236.

Michael, T.P., and McClung, C.R. (2002). Phase-specific circadian clock regulatory elements in *Arabidopsis*. Plant Physiol. **130:** 627–638.

Michael, T.P., and McClung, C.R. (2003). Enhancer trapping reveals widespread circadian clock transcriptional control in *Arabidopsis.* Plant Physiol. **132:** 629–639.

Michael, T.P., et al. (2008). Network discovery pipeline elucidates conserved time-of-day-specific cis-regulatory modules. PLoS Genet. **4:** e14.

Mignone, F., Gissi, C., Liuni, S., and Pesole, G. (2002). Untranslated regions of mRNAs. Genome Biol. **3:** reviews 0004.1–0004.10.

Millar, A.J., Short, S.R., Hiratsuka, K., Chua, N.-H., and Kay, S.A. (1992). Firefly luciferase as a reporter of regulated gene expression in higher plants. Plant Mol. Biol. Rep. **10:** 324–337.

Mizoguchi, T., Wheatley, K., Hanzawa, Y., Wright, L., Mizoguchi, M., Song, H.R., Carre, I.A., and Coupland, G. (2002). *LHY* and *CCA1* are partially redundant genes required to maintain circadian rhythms in *Arabidopsis.* Dev. Cell **2:** 629–641.

Mockler, T.C., Michael, T.P., Priest, H.D., Shen, R., Sullivan, C.M., Givan, S.A., McEntee, C., Kay, S.A., and Chory, J. (2007). The DIURNAL project: DIURNAL and circadian expression profiling, model-based pattern matching, and promoter analysis. Cold Spring Harb. Symp. Quant. Biol. **72:** 353–363.

Molina, C., and Grotewold, E. (2005). Genome wide analysis of *Arabidopsis* core promoters. BMC Genomics **6:** 25.

Oda, A., Fujiwara, S., Kamada, H., Coupland, G., and Mizoguchi, T. (2004). Antisense suppression of the *Arabidopsis PIF3* gene does not affect circadian rhythms but causes early flowering and increases *FT* expression. FEBS Lett. **557:** 259–264.

Plautz, J.D., Straume, M., Stanewsky, R., Jamison, C.F., Brandes, C., Dowse, H.B., Hall, J.C., and Kay, S.A. (1997). Quantitative analysis of *Drosophila period* gene transcription in living animals. J. Biol. Rhythms **12:** 204–217.

Pruneda-Paz, J.L., Breton, G., Para, A., and Kay, S.A. (2009). A functional genomics approach reveals *CHE* as a component of the *Arabidopsis* circadian clock. Science **323:** 1481–1485.

Riechmann, J.L., et al. (2000). *Arabidopsis* transcription factors: Genome-wide comparative analysis among eukaryotes. Science **290:** 2105–2110.

Robertson, M., Helliwell, C.A., and Dennis, E.S. (2008). Post-translational modifications of the endogenous and transgenic FLC protein in *Arabidopsis thaliana.* Plant Cell Physiol. **49:** 1859–1866.

Roden, L.C., Song, H.-R., Jackson, S., Morris, K., and Carré, I.A. (2002). Floral responses to photoperiod are correlated with the timing of rhythmic gene expression relative to dawn and dusk, in *Arabidopsis.* Proc. Natl. Acad. Sci. USA **99:** 13313–13318.

Schaffer, R., Ramsay, N., Samach, A., Corden, S., Putterill, J., Carré, I.A., and Coupland, G. (1998). The *late elongated hypocotyl* mutation of *Arabidopsis* disrupts circadian rhythms and the photoperiodic control of flowering. Cell **93:** 1219–1229.

Shore, P., and Sharrocks, A.D. (1995). The MADS-box family of transcription factors. Eur. J. Biochem. **229:** 1–13.

Suarez-Lopez, P., Wheatley, K., Robson, F., Onouchi, H., Valverde, F., and Coupland, G. (2001). *CONSTANS* mediates between the circadian clock and control of flowering in *Arabidopsis.* Nature **410:** 1116–1120.

Swarup, K., Alonso-Blanco, C., Lynn, J.R., Michaels, S.D., Amasino, R.M., Koornneef, M., and Millar, A.J. (1999). Natural allelic variation identifies new genes in the *Arabidopsis* circadian system. Plant J. **20:** 67–77.

Tanay, A. (2006). Extensive low-affinity transcriptional interactions in the yeast genome. Genome Res. **16:** 962–972.

Viczián, A., Kircher, S., Fejes, E., Millar, A.J., Schäfer, E., Kozma-Bognár, L., and Nagy, F. (2005). Functional characterization of *phytochrome interacting factor 3* for the *Arabidopsis thaliana* circadian clockwork. Plant Cell Physiol. **46:** 1591–1602.

Wang, Z.Y., Kenigsbuch, D., Sun, L., Harel, E., Ong, M.S., and Tobin, E.M. (1997). A Myb-related transcription factor is involved in the phytochrome regulation of an *Arabidopsis Lhcb* gene. Plant Cell **9:** 491–507.

Williams, M.E., Foster, R., and Chua, N.-H. (1992). Sequences flanking the hexameric G-box core CACGTG affect the specificity of protein binding. Plant Cell **4:** 485–496.

Yamashino, T., Matsushika, A., Fujimori, T., Sato, S., Kato, T., Tabata, S., and Mizuno, T. (2003). A Link between circadian-controlled bHLH factors and the APRR1/TOC1 quintet in *Arabidopsis thaliana.* Plant Cell Physiol. **44:** 619–629.

Yanovsky, M.J., and Kay, S.A. (2002). Molecular basis of seasonal time measurement in *Arabidopsis.* Nature **419:** 308–312.

Zeilinger, M.N., Farré, E.M., Taylor, S.R., Kay, S.A., and Doyle, F.J.I. (2006). A novel computational model of the circadian clock in *Arabidopsis* that incorporates PRR7 and PRR9. Mol. Syst. Biol. **2:** 58.