6-2003

Enhancer Trapping Reveals Widespread Circadian Clock Transcriptional Control in Arabidopsis

Todd P. Michael
Dartmouth College

C Robertson McClung
Dartmouth College

Follow this and additional works at: https://digitalcommons.dartmouth.edu/facoa

Part of the Biology Commons

Recommended Citation
Michael, Todd P. and McClung, C Robertson, "Enhancer Trapping Reveals Widespread Circadian Clock Transcriptional Control in Arabidopsis" (2003). Open Dartmouth: Faculty Open Access Articles. 3677.
https://digitalcommons.dartmouth.edu/facoa/3677

This Article is brought to you for free and open access by Dartmouth Digital Commons. It has been accepted for inclusion in Open Dartmouth: Faculty Open Access Articles by an authorized administrator of Dartmouth Digital Commons. For more information, please contact dartmouthdigitalcommons@groups.dartmouth.edu.
Enhancer Trapping Reveals Widespread Circadian Clock Transcriptional Control in Arabidopsis1[w]

Todd P. Michael and C. Robertson McClung*
Department of Biological Sciences, Dartmouth College, Hanover, New Hampshire 03755

The circadian clock synchronizes the internal biology of an organism with the environment and has been shown to be widespread among organisms. Microarray experiments have shown that the circadian clock regulates mRNA abundance of about 10% of the transcriptome in plants, invertebrates, and mammals. In contrast, the circadian clock regulates the transcription of the virtually all cyanobacterial genes. To determine the extent to which the circadian clock controls transcription in Arabidopsis, we used in vivo enhancer trapping. We found that 36% of our enhancer trap lines display circadian-regulated transcription, which is much higher than estimates of circadian regulation based on analysis of steady-state mRNA abundance. Individual lines identified by enhancer trapping exhibit peak transcription rates at circadian phases spanning the complete circadian cycle. Flanking genomic sequence was identified for 23 enhancer trap lines to identify clock-controlled genes (CCG-ETs). Promoter analysis of CCG-ETs failed to predict new circadian clock response elements (CCREs), although previously defined CCREs, the CCA1-binding site, and the evening element were identified. However, many CCGs lack either the CCA1-binding site or the evening element; therefore, the presence of these CCREs is insufficient to confer circadian regulation, and it is clear that additional elements play critical roles.

Timing is everything, making biological clocks central to life in a rhythmic world (Young and Kay, 2001). One such biological clock is the circadian clock, an endogenous oscillator with a period of about 24 h. The circadian clock synchronizes internal biology with the external daily cycle of light and temperature. Biological activities are partitioned, or phased, to specific times of the day, and, therefore, occur in distinct relationships, or phase angles, with each other and the external environment. For example, humans are maximally active during the day, whereas melatonin synthetic rates and blood melatonin levels peak at night (Rutter et al., 2002). In unicellular nitrogen-fixing cyanobacteria, it is critical that oxygenic photosynthesis and anaerobic nitrogen assimilation occur at distinct times of day (Johnson and Golden, 1999). In higher plants, photosynthetic rates peak during the day, whereas hypocotyl and inflorescence stem elongation is maximal at night (McClung et al., 2002).

Rhythmic gene expression underlies many of these complex circadian rhythms. Recent advances in microarray technology have provided the ability in model organisms to ask how many mRNAs oscillate with a circadian period. Such microarray experiments have revealed in mammals, Drosophila melanogaster, Neurospora crassa, and Arabidopsis that between 2% and 10% of mRNAs are regulated by the circadian clock (Harmer et al., 2000; Claridge-Chang et al., 2001; McDonald and Rosbash, 2001; Schaffer et al., 2001; Akhtar et al., 2002; Ceriani et al., 2002; Duffield et al., 2002; Lin et al., 2002; Panda et al., 2002a; Storch et al., 2002; Ueda et al., 2002). These microarray experiments highlight the regulation by the circadian clock of many key biological pathways, including pathways involved in metabolism, detoxification, and stress response. It is possible that the circadian clock regulates a larger portion of biological processes by regulating key steps in pathways rather than entire pathways (Harmer et al., 2000). Although we are accustomed to thinking of “circadian-regulated” genes as those whose transcripts oscillate with a period of 24 h, perhaps this definition should be broadened. Clearly oscillations in mRNA abundance cannot account for all circadian-mediated biology because both posttranscriptional and posttranslational regulation have been shown to play a major role in clock function (Lee et al., 2000, 2001). Moreover, there is potentially another class of genes for which the response to environmental or other stimuli is temporally gated by the circadian clock, and such genes would not be identified in the absence of the appropriate stimulus.

Efforts to understand the molecular mechanisms by which the circadian clock regulates transcription have focused on the identification and characterization of cis-acting elements necessary for circadian-regulated transcription and their cognate DNA-binding proteins. In mammals and D. melanogaster, both the enhancer box (E box; CACGTG) and the cAMP response

1 This work was supported by the National Science Foundation (grant nos. IBN 9817603 and MCB 0091008) and by the U.S. Department of Agriculture (National Research Initiative Competitive Grants Program grant no. 9602632 to C.R.M.).

[w] The online version of this article contains Web-only data. The supplemental material is available at http://www.plantphysiol.org.

* Corresponding author; e-mail mcclung@dartmouth.edu; fax 603–646–1347.

Article, publication date, and citation information can be found at www.plantphysiol.org/cgi/doi/10.1104/pp.021006.
element (TGACGTCA) have been shown to be functional circadian clock response elements (CCREs; Young and Kay, 2001). It is interesting that both the E box and the cAMP response element serve in output genes as the targets of central clock components but are also found in the promoters of genes encoding these same central clock components (Kyriacou and Rosato, 2000). In D. melanogaster, CCREs have been shown to be overrepresented in the predicted promoters of genes whose mRNA oscillates (Claridge-Chang et al., 2001). In contrast, in mice (Mus musculus), the E box was only found in the predicted promoters of a small proportion of cycling genes (Panda et al., 2002a). At least two CCREs have been defined in N. crassa. Some output genes are controlled by the circadian clock through the activating clock element (GTGG-GAT; Bell-Pedersen et al., 1996). In addition, the distal light response element confers light- and circadian-regulated transcription upon the central clock component frequency (Froehlich et al., 2002).

Transcriptional/translational negative feedback loops underpin the molecular mechanisms of the circadian clock (Young and Kay, 2001). Although still incompletely defined, the Arabidopsis circadian clock includes reciprocal regulation between the pseudo response regulator TOC1 (TIMING OF CAB EXPRESSION 1) and two single MYB domain transcription factors, LHY (LATE ELONGATED HYPOCOTYL) and CCA1 (CIRCADIAN CLOCK ASSOCIATED 1; Alabadí et al., 2001, 2002; Matsushika et al., 2002b; Mizoguchi et al., 2002). In plants, functionally defined CCREs are distinct from those identified in mammals and flies. The CCA1-binding site (CBS; AAAATCT) and evening element (EE; AAAATA-TCT) are targets of the central clock components CCA1 and LHY (Wang et al., 1997; Schaffer et al., 1998; Alabadí et al., 2001). In one study, the EE was found to be overrepresented in the predicted promoters of genes with mRNA abundance peaking in the evening (Harmer et al., 2000). Both CCA1 and LHY are induced by light through direct binding of PIF3 (PHOTOTRICHROME INTERACTING FACTOR 3) at the G box (CCACGTGG; Martínez-García et al., 2000). Although the plant G box shares the same core sequence as the animal E box, the factors that bind to these elements are distinct, and it is likely that this sequence similarity is coincidental.

Current efforts to define the subset of the transcriptome that is regulated by the circadian clock have largely relied on transcript profiling, in which large numbers of genes were assayed in parallel to identify transcripts that showed circadian oscillations in abundance. This approach, however, does not directly address circadian regulation of transcription. Transcript stability might obscure oscillations in transcriptional activity, as has been demonstrated for Arabidopsis CAB1 (Millar and Kay, 1991). Random insertion of a transposon-borne promoter-less luciferase reporter gene was used to establish widespread circadian regulation of transcription in cyanobacteria (Liu et al., 1995a). Gene or enhancer trapping (Sundarasan et al., 1995; Campisi et al., 1999; Springer, 2000), in which a reporter gene either lacking a promoter or possessing a minimal unregulated promoter is randomly inserted through T-DNA transformation, has proven effective in Arabidopsis. Therefore, utilizing an in vivo luciferase enhancer trapping technique, we set out to ask to what extent the circadian clock in Arabidopsis regulates transcription. In this study, we reveal that 36% of the Arabidopsis genome is under circadian regulation at the transcriptional level. We identify 23 clock-controlled genes (CCG-ETs) with peak luciferase activity at distinct circadian phases spanning the day. The CBS and EE were found in the predicted promoters of the CCG-ETs identified in the enhancer trap lines.

RESULTS

Enhancer Trapping Circadian-Regulated Elements in Arabidopsis

To identify circadian CCGs and their cognate CCREs and also to establish the extent of transcriptional regulation by the circadian clock in the Arabidopsis genome, we applied an in vivo luciferase-based enhancer trap strategy. Utilizing T-DNA-mediated transformation, a minimal promoter::LUC (LUCIFERASE) construct that was not clock regulated was randomly integrated into the Arabidopsis genome (Michael and McClung, 2002). Individual T1 plants resistant to gentamicin and, hence, carrying the transgene were identified and allowed to self-fertilize. Individual T2 plants resistant to gentamicin were entrained to a light/dark (12/12) cycle and then assayed under continuous light for clock-regulated luciferase activity in vivo using a Packard TopCount Luminometer (Packard, Meriden, CT; Michael and McClung, 2002). Each independent transgenic line exhibiting clock-regulated luciferase activity displayed a characteristic and consistent circadian phase and expression level (Fig. 1; see supplemental data Table III at http://www.plantphysiol.org). Of 335 lines assayed, 128 (36%) were scored as rhythmic by FFT-NLLS (Plautz et al., 1997). Rhythmic lines were defined as having a period of 20 to 30 h and a relative amplitude error (RAE), a measure of the strength of the rhythm, of <1.0. A perfect noise-free cosine wave would return an RAE = 0, whereas an RAE = 1 is the limit of statistical significance for any given rhythmic amplitude. We also generated lines bearing insertions of luciferase lacking a minimal promoter and found the same frequency of rhythmic lines, indicating that the minimal promoter sequences were not necessary for rhythmic transcription of the construct (data not shown). The range of periods (22.41–28.46 h) exhibited by the rhythmic lines is relatively broad. For example, Thain et al. (2002) observed a 2-h difference in the periods of CAB2 and CHALCONE SYN-
to identify lines with peak luciferase activity at circadian phases spanning the circadian day (Fig. 1B; see supplemental data Table III at http://www.plantphysiol.org) and over a broad range of expression levels (Fig. 1C; see supplemental data Table III at www.plantphysiol.org) leads us to believe that we were successful in enhancer trapping CCREs and identifying CCGs.

#### Identification of Enhancer-Trapped Sequences

We defined the T-DNA insertion sites in a subset of 23 circadian-regulated enhancer trap lines by thermal asymmetric interlaced (TAIL)-PCR (Table I; Fig. 2). We chose the 23 enhancer trap lines randomly from the 128 rhythmic lines identified and found that they were evenly distributed over the Arabidopsis genome (Fig. 1D). All 23 of the recovered flanking sequences were aligned to the Arabidopsis sequence database but failed to correspond to genomic, mitochondrial, or chloroplastic DNA, suggesting that it originates from a region of the Arabidopsis genome for which the sequence has not yet been determined (data not shown). Clock-controlled genes as determined by enhancer trapping (CCG-ET) were defined by the orientation of T-DNA insertion relative to predicted Arabidopsis genes: promoter, 5′-UTR, gene (intron or exon), 3′-UTR, and other (Table I; Fig. 2).

Recently, two groups have utilized microarray experiments to identify Arabidopsis mRNAs with a circadian pattern of expression. Six and eight of the CCG-ETs were present on the microarrays used by

---

**Figure 1.** Circadian enhancer trap using luciferase. Seedlings were grown 7 d in 12 h of light and 12 h of dark, and luminescence was measured in continuous light for 5 to 7 d. A, Circadian clock-controlled genes identified by enhancer trapping (CCG-ET) are phased to distinct times of day. Traces were normalized to the average expression over the experiment. All traces are the average of at least 12 seedlings. B, Number of CCG-ET lines with specific circadian phases: circadian time (CT) 0 (0–5.9 h), 6 (6–11.9 h), 12 (12–17.9 h), and 18 (18–23.9 h). All luminescence traces were analyzed for rhythmicity by fast Fourier transform-nonlinear least squares analysis (FFT-NLLS), and circadian phase was determined by normalizing for the period (phase = 24 h × phase/period). C, Expression level of the CCG-ET lines. Expression level was determined directly from the number of photons collected over the sampling interval: 200 (0–199 photons/s), 2,000 (200–1,999 photons/s), 20,000 (2,000–19,999 photons/s), and 200,000 (>20,000 photons/s). All data represent averages from at least two independent experiments. D, Map positions of CCG-ETs on the five Arabidopsis chromosomes.
Table 1. Clock-controlled genes as predicted by enhancer trapping (CCG-ET)

<table>
<thead>
<tr>
<th>Name</th>
<th>Putative Function</th>
<th>Gene Identification</th>
<th>Ins</th>
<th>Expr</th>
<th>Phase</th>
<th>Micro</th>
<th>CBS</th>
<th>EE</th>
<th>G Box</th>
<th>HEX</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCG-ET-1</td>
<td>Elongation factor 1-α</td>
<td>At1g07390</td>
<td>O</td>
<td>200</td>
<td>18</td>
<td>S</td>
<td>1 (10)</td>
<td>0 (8)</td>
<td>0 (0)</td>
<td>0 (2)</td>
</tr>
<tr>
<td>CCG-ET-2</td>
<td>STP1, Glc transporter</td>
<td>At1g11260</td>
<td>O</td>
<td>2,000</td>
<td>8</td>
<td>S, H^g</td>
<td>0 (2)</td>
<td>0 (4)</td>
<td>1 (4)</td>
<td>1 (3)</td>
</tr>
<tr>
<td>CCG-ET-3</td>
<td>SGR2, phospholipase</td>
<td>At1g31480</td>
<td>P</td>
<td>200,000</td>
<td>6</td>
<td>S</td>
<td>0 (10)</td>
<td>0 (7)</td>
<td>0 (0)</td>
<td>0 (1)</td>
</tr>
<tr>
<td>CCG-ET-4</td>
<td>Glyceraldehyde-3-phosphate dehydrogenase</td>
<td>At1g42970</td>
<td>G</td>
<td>2,000</td>
<td>6</td>
<td>H</td>
<td>0 (11)</td>
<td>0 (9)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>CCG-ET-5</td>
<td>Hypothetical, transcription factor/F box</td>
<td>At1g50980</td>
<td>G</td>
<td>200,000</td>
<td>12</td>
<td>ND</td>
<td>1 (11)</td>
<td>1 (3)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>CCG-ET-6</td>
<td>Unknown, MYB, kelch domains</td>
<td>At1g511540</td>
<td>P</td>
<td>20,000</td>
<td>20</td>
<td>ND</td>
<td>1 (6)</td>
<td>0 (6)</td>
<td>0 (2)</td>
<td>0 (2)</td>
</tr>
<tr>
<td>CCG-ET-7</td>
<td>RBP10-α putative RNA polymerase</td>
<td>At1g53690</td>
<td>P</td>
<td>200</td>
<td>12</td>
<td>ND</td>
<td>1 (6)</td>
<td>0 (4)</td>
<td>0 (0)</td>
<td>0 (1)</td>
</tr>
<tr>
<td>CCG-ET-8</td>
<td>Putative, GTP-binding protein</td>
<td>At1g60530</td>
<td>P</td>
<td>200</td>
<td>0</td>
<td>H^b</td>
<td>0 (4)</td>
<td>0 (8)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>CCG-ET-9</td>
<td>Hypothetical protein</td>
<td>At1g73790</td>
<td>G</td>
<td>2,000</td>
<td>8</td>
<td>ND</td>
<td>2 (9)</td>
<td>0 (10)</td>
<td>0 (2)</td>
<td>0 (3)</td>
</tr>
<tr>
<td>CCG-ET-10</td>
<td>Putative, Ser carboxypeptidase I</td>
<td>At2g22980</td>
<td>G</td>
<td>20,000</td>
<td>10</td>
<td>H^b</td>
<td>0 (6)</td>
<td>1 (9)</td>
<td>0 (4)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>CCG-ET-11</td>
<td>Putative, MYB domain</td>
<td>At2g266950</td>
<td>3</td>
<td>2,000</td>
<td>3</td>
<td>H^b</td>
<td>1 (15)</td>
<td>1 (11)</td>
<td>0 (0)</td>
<td>0 (1)</td>
</tr>
<tr>
<td>CCG-ET-12</td>
<td>Putative oxysterol binding</td>
<td>At2g31030</td>
<td>S</td>
<td>2,000</td>
<td>6</td>
<td>ND</td>
<td>1 (7)</td>
<td>0 (8)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>CCG-ET-13</td>
<td>AtKUP3, K^+ transporter</td>
<td>At3g02050</td>
<td>G</td>
<td>2,000</td>
<td>0</td>
<td>S</td>
<td>0 (8)</td>
<td>0 (6)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>CCG-ET-14</td>
<td>Unknown protein</td>
<td>At3g10640</td>
<td>P</td>
<td>200</td>
<td>12</td>
<td>S</td>
<td>0 (13)</td>
<td>1 (5)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>CCG-ET-15</td>
<td>Putative DNA binding</td>
<td>At3g15590</td>
<td>P</td>
<td>2,000</td>
<td>12</td>
<td>ND</td>
<td>1 (9)</td>
<td>0 (4)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>CCG-ET-16</td>
<td>Putative protein</td>
<td>At3g44410</td>
<td>P</td>
<td>2,000</td>
<td>12</td>
<td>S</td>
<td>0 (1)</td>
<td>0 (1)</td>
<td>0 (0)</td>
<td>0 (1)</td>
</tr>
<tr>
<td>CCG-ET-17</td>
<td>Expressed protein</td>
<td>At4g14710</td>
<td>G</td>
<td>2,000</td>
<td>11</td>
<td>S</td>
<td>0 (6)</td>
<td>1 (5)</td>
<td>0 (0)</td>
<td>0 (2)</td>
</tr>
<tr>
<td>CCG-ET-18</td>
<td>Putative, chloroplast DNA binding</td>
<td>At4g21430</td>
<td>P</td>
<td>20,000</td>
<td>12</td>
<td>ND</td>
<td>0 (8)</td>
<td>0 (8)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>CCG-ET-19</td>
<td>Putative, diacylglycerol kinase iota</td>
<td>At4g30340</td>
<td>P</td>
<td>20,000</td>
<td>12</td>
<td>ND</td>
<td>0 (7)</td>
<td>0 (6)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>CCG-ET-20</td>
<td>Putative protein</td>
<td>At5g40270</td>
<td>G</td>
<td>2,000</td>
<td>12</td>
<td>S</td>
<td>0 (8)</td>
<td>0 (5)</td>
<td>0 (2)</td>
<td>0 (2)</td>
</tr>
<tr>
<td>CCG-ET-21</td>
<td>Unknown protein</td>
<td>At5g37170</td>
<td>P</td>
<td>20,000</td>
<td>11</td>
<td>ND</td>
<td>0 (7)</td>
<td>0 (5)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>CCG-ET-22</td>
<td>AtPME3-like, pectin methylesterase</td>
<td>At5g55590</td>
<td>P</td>
<td>2,000</td>
<td>0</td>
<td>ND</td>
<td>0 (8)</td>
<td>0 (3)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>CCG-ET-23</td>
<td>Putative, mRNA process (splicing)</td>
<td>At5g56900</td>
<td>G</td>
<td>200</td>
<td>4</td>
<td>ND</td>
<td>0 (7)</td>
<td>1 (2)</td>
<td>0 (0)</td>
<td>0 (1)</td>
</tr>
</tbody>
</table>

a The Arabidopsis Information Resource and the Munich Information Center for Protein Sequences. b T-DNA insertion position in reference to predicted gene location. P, promoter (−2,000/−150); G, exon or intron; 5′-UTR (−150/−1 bp); 3′-UTR; O, other. Average bioluminescence (photons per second per seedling). d Phase, Time of the peak luciferase activity in continuous conditions normalized to CT. e Micro, Microarray experiments; S, Schaffer et al. (2001); H, Harmer et al. (2000). f CBS (AAAAATTCA), EE (AAATATCT), G box (CCACGTGG), and HEX (TGACGTGG); no. of exact matches in predicted promoters (one mismatch). g Rhythmically expressed. h ND, Not determined.

Harmer et al. (2000) and Schaffer et al. (2001), respectively. CCG-ET2, encoding the STP1 Glc transporter (At1g11260), was determined to be circadian regulated by Harmer et al. (2000), whereas Schaffer et al. (2001) characterized it as diurnally regulated, yet not circadian regulated (Table I). CCG-ET4, encoding glyceraldehyde-3-phosphate dehydrogenase (At1g42970), was not found to be rhythmic in continuous light by Harmer et al. (2000), yet is circadian regulated in other systems (Lillo, 1993; Shinohara et al., 1998; Fagan et al., 1999). The four remaining CCG-ETs represented on the microarray tested by Harmer et al. (2000) displayed no detectable signal (Table I; S. Harmer and S.A. Kay, personal communication), and so neither confirm nor refute their regulation by the circadian clock. Although CCG-ET22 (AtPME3-like, At5g55590; Table I) was not on either of the microarrays, its putative paralog, AtPME3, was characterized as rhythmic by Harmer et al. (2000).

CCREs

Enhancer trapping takes an unbiased approach to the identification of genes whose transcript is clock-regulated and, therefore, may reveal novel CCREs. To identify possible CCREs, we searched the predicted promoters (defined as −2,000/−1 upstream of the predicted translational start site) of all the CCG-ETs for overrepresented elements using the Web-based regulatory sequence analysis tools (RSAs; http://embnet.cifn.unam.mx/rsa-tools/). Only one element between 5 and 8 bp in length, GATA/CAC, was found to be significantly overrepresented. It was only present in 39% (9 of 23) of the CCG-ET promoters. The promoter region of CCG-ET15 (At5g15950) contained 17 of these motifs, accounting for 56% (17 of 30) of the motifs found in this set of 23 promoters. When CCG-ET15 is removed from the analysis, there are no significantly overrepresented elements in the promoters of the CCG-ETs. Repeating the analysis on subsets of the CCG-ETs grouped according to the circadian phase of peak expression did not reveal overrepresented elements.

A similar promoter alignment tool (AlignAce; http://arep.med.harvard.edu/) identified the EE (AAATATCT) in the promoters of the evening-specific (CT 8 and CT 12, where CT 0 is subjective dawn) CCGs identified by Harmer et al. (2000). Utilizing RSA, we also identified the EE (AAATA/TCTA and AAAATATCTT) in promoters of rhythmically genes with evening-specific phases. In genes with morning-specific phases (CT 20 and CT 0), the G-box core (CACGTG) was overrepresented with the consensus (GACACGTGG). G boxes located in the CA1 and LHY promoters serve as the site of activation by the light-transducing transcription factor.
PIF3 (Martínez-García et al., 2000). Therefore, like the light response element in N. crassa that confers both light and circadian regulation (Froehlich et al., 2002), the G box may also play dual roles as has been hypothesized (Menkens et al., 1995).

We also used RSA to search the predicted promoters of genes thought to be intimately involved in the circadian clock. When the promoters of CCA1, LHY, TOC1 (APRR1), APRR3, APRR5, APRR7, and APRR9 (Matsushika et al., 2002a; McClung et al., 2002; Sato et al., 2002) were searched together, three oligos were overrepresented: EE (AAATATCT), G box (CACGTCGTC) and the Hex element (TGACGTGG; Schindler et al., 1992; Fig. 3). The Hex element is related to the G box and is bound by G-box-binding factors (Schindler et al., 1992). The Hex element is overrepresented specifically in the promoters of the family of Arabidopsis APRRs (PSEUDO-RESPONSE REGULATORS), of which the central clock component TOC1 is the founding member (Fig. 3D; Matsushika et al., 2000; Strayer et al., 2000).

Promoter Analysis of the Arabidopsis Genome

To place the CBS, EE, G box, and Hex element results from the CCG-ETs in the context of the whole Arabidopsis genome, we searched for these motifs in promoters (−2,000/−1 from the ATG) from most (25,545) of the predicted Arabidopsis genes. To determine the “expected” frequency of these motifs, we also randomly generated 25,545 “promoters,” each 2,000 bp, which were derived with the same base pair composition as actual Arabidopsis promoters (see “Materials and Methods”). We also analyzed the predicted promoters from 8,200 genes on the Affymetrix GeneChip, the Harmer et al. (2000) clock-regulated genes, the Arabidopsis Functional Genomics Consortium (AFGC) microarray genes, and the genes described by Schaffer et al. (2001) as clock regulated, diurnally regulated, light induced, or dark induced. Figure 4 illustrates the occurrence of the CBS, EE, and G box in the predicted promoters of these sets of genes, and Table II summarizes our findings.

Of the four motifs sampled, the two known CCREs are found slightly less frequently (approximately 0.7-fold) in the Arabidopsis promoter database than predicted by chance, whereas the G-box and Hex elements are found more frequently (1.96- and 1.56-fold, respectively). The frequencies of these elements in Figure 2. Insertions of LUCIFERASE relative to predicted Arabidopsis genes in the enhancer trap lines. T-DNA carrying the LUCIFERASE gene (shaded arrow) inserted as indicated into Arabidopsis genomic sequence in relation to the predicted CCG-ET (black arrow) and the nearest adjacent gene (white arrow). The distance (base pairs) from the ATG (or STOP) of the predicted CCG-ET to the start of the LUCIFERASE gene is indicated. Also indicated, below the horizontal line of the chromosome, is the distance between the predicted CCG-ET and the nearest adjacent gene. Depictions are not to scale.
the promoters of genes on both microarrays are representative of their frequency in the whole genome, except for a modest enrichment of the G box (Table II). A number of observations can be made. First, the G-Box is overrepresented in the predicted promoters of three sets of clock-regulated genes (Table II; CCG-ET, Harmer circadian and Schaffer circadian) and of genes that are diurnally and light regulated (Table II; Schaffer light induced and Schaffer diurnal). Consistent with the G box playing a role in light signaling, it is overrepresented among light-induced genes and underrepresented among dark-induced genes (Table II) (Schaffer et al., 2001). Second, the EE is slightly overrepresented in the Harmer circadian, Schaffer circadian, Schaffer dark-induced, and Schaffer diurnal sets of genes (Table II). Third, the Hex element is overrepresented in the Schaffer light-induced, dark-induced, and diurnal gene sets, and among the Harmer circadian genes (Table II). Finally, the occurrence of the CBS does not vary among the sets of genes (Table II).

The CBS and EE are found in the promoters of 35% and 25%, respectively, of Arabidopsis genes, which is consistent with our estimation that 36% of the Arabidopsis genome is circadian regulated. Is the presence of a CBS or EE sufficient to predict clock regulation? Apparently not at the level of mRNA abundance, because the CBS and EE are found in 35% and 25%, respectively, of the promoters on both the Affymetrix and AFGC microarrays, yet only 6% and 2%, respectively, of all the genes on these microarrays were found to be clock regulated. In fact, 93% (3,236 of 3,471) and 97% (3,384 of 3,487) of the genes on the Affymetrix and AFGC, respectively, whose predicted promoters have the CBS or EE, were not identified as rhythmic. Clearly, the presence of a CBS or EE is insufficient to confer circadian-regulated mRNA abundance. This is consistent with the accumulating evidence that additional elements are required to provide a context in which CCREs can properly function (Michael and McClung, 2002; Muñoz et al., 2002).

Combinatorial interaction of light-responsive elements confers light-regulated expression in Arabidopsis (Puente et al., 1996; Chattopadhyay et al., 1998). Clock regulation may require multiple copies or combinations of the CBS, EE, and/or G box. Consistent with this idea, 76% and 83% of the G boxes identified in the promoters of Harmer et al. (2000) and Schaffer et al. (2001) CCGs, respectively, occurred in combination with either the CBS or EE (Fig. 4, E and F), which is an enrichment relative to the whole-genome frequency of co-occurrence of approximately 50% (Fig. 4A, B, C). Among the CCGs identified by Harmer et al. (2000), the presence of three or more EEs in a promoter is a stronger predictor of rhythmicity than the presence of a single element, but only one-third of the genes with three or more copies of the EE are rhythmic, whereas two-thirds are not. In

![Figure 3. CCREs in the promoters of genes encoding clock components. Predicted promoters (-2,000/-1 upstream of the ATG) of CCA1, LHY, APRR9/TOC1, APRR3, APRR5, APRR7, and APRR9 were aligned using AlignACE (http://arep.med.harvard.edu; Hughes et al., 2000) and RSA (http://embnet.cifn.unam.mx/rsa-tools/). Motif models were created as sequence logos (Schneider and Stephens, 1990) using the program GENIO/logo (http://genio.informatik.uni-stuttgart.de/GENIO/logo.cgi). At each position, the height of each residue is proportional to its frequency in that position. The total height of all the residues in the position is proportional to the degree of conservation (information content) of the position, presented in bits. Two bits are sufficient to specify a single nucleotide as always present at a given position and, therefore, are the maximum information content possible at any position. Upside-down letters appear if the probability of that nucleotide occupying the position is less than the equally distributed a priori probability of all letters. A, CBS/EE (AAAa/tATCT); B, G box (CACGTG); C, HEX element (TGACGTGG); D, Number of each element and total number of elements in each predicted promoter.](http://arep.med.harvard.edu)
contrast, the presence of three or more copies of the CBS did not correlate with an increased frequency of rhythmicity (Table II).

DISCUSSION

Our observation that 36% of the Arabidopsis genome is potentially under transcriptional control by the circadian clock suggests a much more pervasive role for the clock in the regulation of gene expression than do estimates generated through microarray analysis of mRNA oscillations, which suggest that 2% to 6% of the transcriptome is clock regulated (Harmer et al., 2000; Schaffer et al., 2001). In D. melanogaster, microarray analysis and enhancer trapping provide similar estimates of the degree (approximately 10%) of clock control of gene expression (Claridge-Chang et al., 2001; McDonald and Rosbash, 2001; Stempfl et al., 2002; Ueda et al., 2002), and these estimates are largely congruent with estimates from microarray analyses conducted in mammals (Grundschober et al., 2001; Akhtar et al., 2002; Duffield et al., 2002; Panda et al., 2002a; Storch et al., 2002). However, promoter trap experiments in cyanobacteria have revealed that transcription of most genes is regulated by the circadian clock (Liu et al., 1995a; Aoki et al., 2002).

Possibly the most parsimonious way to explain the difference between the transcriptional and mRNA abundance estimates of circadian regulation is to recall that mRNA abundance does not strictly reflect transcriptional activity. For oscillations in transcription to yield oscillations in mRNA abundance requires that the transcript be sufficiently unstable to turn over within a circadian cycle. For example, it has been demonstrated that both oscillations in de novo transcription rates and in mRNA stability contribute to oscillations in mRNA abundance for chloroplast genes in Chlamydomonas reinhardtii (Salvador et al., 1993; Hwang et al., 1996). A recent estimate is that the majority of Arabidopsis transcripts (>95%) have half-lives >2 h (Gutierrez et al., 2002), suggesting that there may be a subset of genes for which the transcript stability is too great to allow clock-regulated changes in de novo transcription to be apparent as mRNA abundance oscillations. For instance, circadian oscillations in transcriptional activity of the CATALASE 3 (CAT3) promoter in light-dark cycles or in continuous light result in robust oscillations in CAT3 transcript abundance (Zhong and McClung, 1996; Michael and McClung, 2002). However, in continuous dark, the robust circadian oscillations in transcriptional activity persist, but the CAT3 transcript accumulates to a high and nonoscillating level, presumably due to an increase in transcript stability (Zhong et al., 1997; Michael and McClung, 2002). Similarly, although the majority of genes of Synechococcus elongatus PCC7942 exhibit robust circadian oscillations in transcription, as measured by activity of luciferase gene fusions (Liu et al., 1995a), parallel oscillations are not always detected at the level of mRNA abundance by northern-blot analysis (S.S. Golden, personal communication).

Other factors may prevent the detection of circadian oscillations in transcription at the level of microarray analysis. Sampling of total mRNA may mask tissue-specific oscillations of genes that only cycle in a subset of the tissues in which they are expressed. Enhancer trapping does not bias against tissue-specific expression, although it should be noted that our activity measure is primarily from the leaves (Thain et al., 2002). The luciferase assay is extremely sensitive, allowing the reproducible detection of oscillations at or below the limits of detection in a standard mRNA-based assay. This sensitivity is enhanced by our ability to monitor rhythms of multiple seedlings for each insertion for 5 to 7 d, taking measurements every hour. In addition, we can retest specific lines in multiple experiments. Thus, our as-

---

**Figure 4.** Occurrence of CBS, EE, and G box in predicted Arabidopsis promoters. Predicted promoters (~2,000/−1 upstream of the ATG) for most Arabidopsis genes (25,545) were searched for the following motifs: CBS (AAAAATCT), EE (AAATATCT), G box (CCACGTGG), and G box core (CACGTG). Venn diagrams illustrate the occurrences of the CBS (upper left), EE (upper right), and G box (bottom). A, Genes from the Arabidopsis genome; B, genes on the Affymetrix Gene Chip (AffyGC); C, genes on the AFGC microarray; D, CCG-ETs; E, clock-controlled genes detected by Harmer et al. (2000); F, clock-controlled genes detected by Schaffer et al. (2001).
say allows us to reliably identify low-amplitude rhythms or weakly expressed CCGs (Fig. 1C). Such genes are very likely filtered out by the stringent statistical criteria used in microarray analysis to reduce the identification of false positives.

Our prediction of the enhancer-trapped CCG-ETs is based on proximity and orientation of the inserted LUC gene relative to the start site of the putative CCG-ET. In each case, we chose the gene whose start is closest to the LUC start except when LUC had inserted downstream of the start and in the antisense orientation. These criteria may have allowed the misidentification of some CCG-ETs. Targeted transcriptional LUC fusions to each of the 46 genes adjacent to the sites of LUC insertion in the 23 characterized lines would allow unambiguous resolution of this issue. It is also possible that we have identified “orphan” CCREs that fail to confer circadian regulation on adjacent genes but are capable of conferring circadian regulation on our LUC construct. Although plant regulatory regions tend to be compact, it is also possible that CCREs may be able to act at a distance

(Bertolino and Singh, 2002). Therefore, circadian regulation may have been incorrectly attributed to a gene that is not regulated itself but that resides adjacent to a gene that is regulated by the clock. In any case, we do not think that the incorrect prediction of some of the CCG-ETs undermines the conclusion that there may be more circadian regulation of transcription than is indicated by analysis of mRNA abundance.

Although our CCG-ETs are distributed evenly over the chromosomes, T-DNA integration is not completely random. Two studies have suggested a slight bias toward integration into intergenic regions of high A + T base composition (Brunaud et al., 2002; Krysan et al., 2002). This should serve to increase insertion near promoter elements and, hence, overestimate the frequency of regulatory elements capable of conferring circadian regulation. However, in both studies, this bias was modest and insufficient to account for the much higher estimate of clock-controlled transcription (36%) versus mRNA abundance oscillations (6%–10%). Rhythmic histone acetylation recently has been correlated with circadian

Table II. Occurrence of CBS, EE, G-Box, and Hex elements in predicted Arabidopsis promoters

<table>
<thead>
<tr>
<th>Motif 8mer</th>
<th>Observed % (i)</th>
<th>Expected % (i)</th>
<th>Obs/exp</th>
<th>Observed % (ii)</th>
<th>Expected % (ii)</th>
<th>Obs/exp</th>
<th>Observed % (iii)</th>
<th>Expected % (iii)</th>
<th>Obs/exp</th>
</tr>
</thead>
<tbody>
<tr>
<td>CBS AAAATCT</td>
<td>35 (9035)</td>
<td>53 (13641)</td>
<td>0.66</td>
<td>36 (2602)</td>
<td>53 (3798)</td>
<td>0.69</td>
<td>35 (2468)</td>
<td>53 (3692)</td>
<td>0.67</td>
</tr>
<tr>
<td>EE AAATATCT</td>
<td>25 (6455)</td>
<td>35 (9042)</td>
<td>0.71</td>
<td>25 (1825)</td>
<td>35 (2508)</td>
<td>0.73</td>
<td>26 (1840)</td>
<td>35 (2508)</td>
<td>0.73</td>
</tr>
<tr>
<td>G-Box CCACGTGG</td>
<td>3 (651)</td>
<td>1 (332)</td>
<td>1.96</td>
<td>3 (188)</td>
<td>1 (71)</td>
<td>2.64</td>
<td>3 (188)</td>
<td>1 (71)</td>
<td>2.64</td>
</tr>
<tr>
<td>Hex TGACGTGG</td>
<td>5 (1274)</td>
<td>3 (817)</td>
<td>1.56</td>
<td>6 (406)</td>
<td>3 (214)</td>
<td>1.89</td>
<td>6 (389)</td>
<td>3 (214)</td>
<td>1.80</td>
</tr>
<tr>
<td>Total Genes</td>
<td>25,545</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Expected values were determined by establishing the frequency of the oligonucleotide in the 25,545 random Arabidopsis promoters. Gene sets are defined as follows: Harmer Circadian, genes defined as circadian regulated in continuous light by Harmer et al. (2000); Schaffer Circadian, genes defined as circadian regulated in continuous light by Schaffer et al. (2001); Schaffer Light Induced, the subset of circadian genes defined Schaffer et al. (2001) in which expression was higher in the day than in the night; Schaffer Dark Induced, the subset of circadian genes defined Schaffer et al. (2001) in which expression was higher in the night than in the day; Schaffer Diurnal, genes shown to oscillate in a light-dark cycle by Schaffer et al. (2001).
rhythms in mRNA abundance in mice, consistent with a key role of chromatin remodeling in circadian transcription (Etchegaray et al., 2003). It has also been suggested that clock-mediated condensation and/or supercoiling of the cyanobacterial chromosome is responsible for global circadian oscillations of transcript abundance (Mori and Johnson, 2001; Mori et al., 2002), similar to that found in the chloroplast of C. reinhardtii (Salvador et al., 1998). Possibly, T-DNA insertions could be biased toward chromosomal regions whose chromatin structure is under circadian control. Clustering of clock-regulated genes has been observed in D. melanogaster (Ueda et al., 2002). In Arabidopsis, the identification of TEl has implicated polyADP-ribosylation in circadian regulation, possibly via chromatin modification (Panda et al., 2002b). DET1 (DEETIOLATED 1) associates with the core histone H2B in the context of the nucleosome, suggesting that chromatin remodeling plays a role in photomorphogenesis (Benvenuto et al., 2002). The det1 mutation also results in circadian defects (Millar et al., 1995), consistent with a role of chromatin structure in circadian regulation of transcription.

Our study suggests that transcriptional control of the Arabidopsis circadian clock may be more widespread than previously demonstrated through microarray experiments. The discrepancies between estimates of clock regulation based on transcription versus mRNA abundance, together with the observation that clock-regulated transcripts are enriched in the set of relatively unstable transcripts (Gutierrez et al., 2002), suggests that mRNA stability may be an important component of circadian regulation of gene expression. High levels of mRNA stability may reduce the effective number of cycling gene products by obscuring oscillations in abundance of transcripts whose transcription is circadian regulated. Conversely, circadian regulation of transcript stability, either periodically stabilizing or periodically destabilizing a transcript, could confer oscillations in mRNA abundance on transcripts whose transcription rates do not vary over the circadian cycle.

MATERIALS AND METHODS

Circadian Enhancer Trap Lines

A modified pLZP plant transformation vector containing a minimal CAT3 promoter (80 bp, insufficient to drive circadian transcription [Michael and McClung, 2002]), an omega translational enhancer, the luciferase coding sequence (LUC+), and the nopaline synthase terminator was used for circadian enhancer trapping. Enhancer trap lines were created using floral dip transformation (Bent and Clough, 1998) of different Arabidopsis ecotypes (Columbia, CS933; Rchew, CS913; Wassilewskija, CS915; and Landsberg erecta, CS20). Agrobacterium tumefaciens strain GV3101 was used for all transformations. T1 seedlings were determined to be rhythmic if their period was between 20 and 30 h, the peak signal strength exceeded 100 photons s−1 seedling−1 and the RAE, a measure of the strength of the rhythm, was <1.0, the limit of statistical significance for any given rhythmic amplitude (Michael and McClung, 2002). For all experiments, between 12 and 24 independent T2 lines were tested in a minimum of two independent experiments. Phase is given in CT (phase = 24 h × phase/period) to facilitate comparison of phase between rhythmic traces with different periods.

Characterization of Enhancer Trap Lines

To characterize possible enhancer-trapped promoters GENES/TAIL-PCR (Liu et al., 1995b) was used to clone genomic Arabidopsis DNA from enhancer trap lines. Genomic DNA was isolated using the Nucleon Phytopure DNA extraction system (Amersham Biosciences, Piscataway, NJ). Vector-specific (pLZP) primers to both the right and left borders were utilized (Hajdúekiewicz et al., 1994). Primary TAIL reactions were subjected to secondary PCR and directly sequenced. The cloning sequence was aligned to the Arabidopsis genome using BLAST.

CCRE Mapping in Arabidopsis Promoters

BSAs (van Helden et al., 2000) were utilized for extracting Arabidopsis promoters, searching for overrepresented oligonucleotides, and searching for CCREs. Arabidopsis promoters were defined as between −2,000 and −1 bp relative to the predicted translational start. Random Arabidopsis promoters (25,545, each 2,000 bp) were generated using the Markov chain process based on nucleotide frequencies of the complete set of Arabidopsis intergenic sequences with a oligonucleotide size of six (http://rsat.ub.edu/rsat/). Expected values were determined by establishing the frequency of the oligonucleotide in the 25,545 random Arabidopsis promoters. Gene sets are defined as follows: Harmer circadian, genes defined as circadian regulated in continuous light by Harmer et al. (2000); Cassaer circadian, genes defined as circadian regulated in continuous light by Schafer et al. (2001) in which expression was higher in the day than in the night; dark induced, the subset of circadian genes defined by Schafer et al. (2001) in which expression was higher in the night than in the day; and Schafer diurnal, genes known to oscillate in a light-dark cycle by Schafer et al. (2001). Genes and motifs were mapped to Arabidopsis chromosomes and visualized using GeneSpring (Silicon Genetics, Redwood City, CA).

ACKNOWLEDGMENTS

We thank Chris Langmead and Bob Gross for Arabidopsis promoter databases and Francisa Cerda for technical assistance with TAIL-PCR. We thank Mary Lou Guerinot and Patrice Salomez for helpful discussions and two anonymous reviewers for useful suggestions. We thank the Arabidopsis Biological Resource Center for all Arabidopsis accessions used in this study.
LITERATURE CITED


Circadian Enhancer Trapping in Arabidopsis


