The Circadian Clock-controlled Transcriptome of Developing Soybean Seeds

Karen A. Hudson*

Abstract
A number of metabolic and physiological processes in plants are controlled by the circadian clock, which enables a plant to anticipate daily changes in the environment. Relatively little is known about circadian rhythms in developing seeds, which may be important for determining the extent and timing of nutrient storage in grain. Microarray expression profiling was used to identify genes expressed in developing soybean (Glycine max) seeds that are controlled by the circadian clock. Genes with predicted functions in protein synthesis, fatty acid metabolism, and photosynthesis totaling 1.8% of the mRNAs detected in seed were found to be expressed in a circadian rhythm. Known circadian and light-controlled promoter elements were identified as over-represented in the promoters of clock-controlled seed genes, with the over-represented elements varying according to the phase of circadian expression. A subset of circadian-regulated genes were found to be expressed in different phases in developing seeds with respect to leaves from the same plants, many of which have roles in photosynthesis and carbon metabolism. These results help to characterize the genes and processes in seeds that may be regulated by the circadian clock, and provide some insight into organ-specific phasing of clock controlled gene expression.

SOYBEAN (Glycine max) is a seed crop rich in protein and oil; consequently the genes involved in soybean seed development and composition have been an area of intense investigation for a number of years (Weber et al., 2005). However, many of the molecular and physiological factors that determine the carbohydrate, protein, and oil content of mature soybean seeds are still incompletely understood. Seed composition is controlled by the metabolic activities within the seed itself as well as maternal factors. Carbon in the form of sugar is transported to the seed and used in the synthesis of complex carbohydrates, storage proteins, and fatty acids (Hills, 2004). In developing soybean seeds, growth and storage compound biosynthesis are both influenced by light (Ruuska et al., 2004; Allen et al., 2009). In seeds, biosynthesis of oils begins in the chloroplast, where gene expression and metabolic state is highly dependent on the light environment (Ohlrogge and Jaworski, 1997). Recently, it has been shown that 5 to 30 μmol m⁻² s⁻¹ light penetrates the soybean pod wall to reach the developing seeds, that seeds undergo photosynthesis, and that light levels within the pod are positively correlated with seed growth. Energy derived from seed photosynthesis provides ATP, reducing power and some carbohydrate in seeds, and contributes to the carbon in storage fatty acids (Allen et al., 2009). Seed photosynthesis has also been shown to increase oxygen content within the pod where gas exchange is limited, and is correlated with starch synthesis (Rolletschek et al., 2003). In intact seedlings it has been shown that numerous genes involved in starch metabolism, photosynthetic light harvesting, and phenylpropanoid biosynthesis are regulated by the circadian clock, which enables the plant to synchronize the occurrence of these processes with the environment.
time of day (Harmer et al., 2000). It has also been shown that genes involved in protein biosynthesis share a specific cis-regulatory motif that confers circadian expression, and these genes are phased so that their expression precedes the daily growth period (Michael et al., 2008a). Clearly, the processes of photosynthesis, plastid development, and carbohydrate, protein, and lipid metabolism are likely to have an important impact on seed development. It has been shown that the Opaque2 gene in developing maize kernels and genes encoding the starch branching enzymes in sorghum endosperm are expressed with a diurnal rhythm (Ciceri et al., 1999; Mutsiya et al., 2009). However, transcriptome profiling techniques have not been applied to investigate circadian rhythms in gene expression in developing seed tissues. Using circadian clock-regulated gene expression to dissect the control of the pathways for oil, protein, and carbohydrate biosynthesis in seeds may provide an opportunity to further understand and influence both seed and yield composition.

The current model of the plant circadian clock consists of three interlocked feedback loops (Locke et al., 2005; McClung, 2006). The soybean circadian oscillator has not been extensively characterized at the molecular and genetic level, but orthologs for many of the clock genes such as LATE ELONGATED HYOCOTYL (LHY) and TIMING OF CAB1 (TOC1) can be found in the soybean genome sequence and have been demonstrated to cycle in a manner similar to that in Arabidopsis (Sullivan et al., 2004; Liu et al., 2008b). Soybean accessions fall into one of a number of maturity groups, which are optimized to flower and produce seed at different latitudes (Morse et al., 1947). There is substantial natural diversity for the maturity trait, and the genes involved are likely to be a source of genetic variation for circadian and photo-period sensing parameters. For example, cryptochrome protein levels vary with latitude of adaptation in Chinese soybean accessions (Zhang et al., 2008). Two paralogous loci encoding phytochrome A-like photoreceptors in soybean have been implicated in photoperiod response and the control of flowering time at the E3 and E4 maturity loci (Liu et al., 2008a; Watanabe et al., 2009). However, a number of other loci affecting maturity remain uncharacterized at the molecular level (Bernard, 1971; Buzzell, 1971; Buzzell and Voldeng, 1980; McBlain and Bernard, 1987; Cober and Voldeng, 2001). Further work is needed to determine how the photoperiod genes may influence yield and other traits such as seed composition.

Organ-specific measurement of circadian reporter genes in tobacco cotyledons has shown that the same gene can be entrained to different phases in the same plant, and therefore the circadian clock can function in a cell- and organ-autonomous manner (Thain et al., 2000). Recent experiments profiling circadian clock-controlled gene expression in Arabidopsis roots found that CIRCADIAN-CLOCK ASSOCIATED1 (CCA1) and LHY oscillated in roots, but TOC1 and other “evening genes” did not. Oscillating expression patterns of some root transcripts were not precisely in phase with their expression in shoots in constant conditions, and it was concluded that the root and shoot clock are synchronized in diurnal cycles by a carbohydrate-related signal (James et al., 2008). This result underscores the possibility that the oscillator and circadian output pathways may be regulated in an organ-specific manner, and the clock may be entrained by transported signals from other organs as well as via temperature and light signals acting cell autonomously.

Many of the genes expressed in developing seeds are not expressed elsewhere in the adult plant, and thus are unlikely to be identified in circadian profiling experiments from whole seedlings and leaf tissue. The size of the soybean seed and its high mRNA content make it a good model for the study of gene expression during seed development, since it is relatively straightforward to extract enough mRNA for many procedures from a small number of seeds (Le et al., 2007). The aim of this study was to determine the extent to which the expression of genes in developing soybean seeds is governed by the circadian clock, and to characterize the rhythmicity and phasing of genes involved in development, seed filling, and composition. A thorough understanding of these processes is necessary before systems biology approaches can be applied to the improvement of seed composition and yield.

MATERIALS AND METHODS

Plant Growth Conditions, RNA Extraction

Four soybean seeds (cv.Williams-82) were planted in one part Promix (BRK): two parts sand/soil mix in 3-gallon pots. Plants were grown in growth chambers (Controlled Environments, Ltd., Winnipeg, Canada) at 25°C and 75% relative humidity under 12 hour light/12 hour dark cycles, with lighting provided by a combination of incandescent and fluorescent bulbs at an intensity of 350 μmol m⁻² s⁻¹. To reduce the effects of seed age on gene expression, flowers were marked as they opened for a total of five consecutive days, and only the marked pods were used for RNA extraction. Five weeks after flower opening, lights were switched on for the rest of the experiment. Marked pods and leaves were collected into liquid nitrogen at 4-h intervals. RNA was extracted from eight seeds or three trifoliate leaflets per sample using a modified version of the “Pine Tree method” from Tepperman et al. (2001). Two RNA samples, each a pool of eight developing seeds from multiple plants, were used for microarray analysis. An additional pool of eight seeds was also collected and used along with the two pools used for microarrays to give an additional replicate for qPCR (quantitative real time polymerase chain reaction). Pooled leaf RNA samples (two pools consisting of one leaflet from each of three individuals), were collected at the same time. Triplicate leaf and seed samples staged in the same manner were taken from additional plants at 4-h intervals grown in diurnal cycles for the 24-h diurnal timecourse measurements.
Microarray Analysis

Labeling and microarray hybridization was performed at the Purdue Genomics Center according to the manufacturers protocols. Microarray data was analyzed using R (2.7.0) and Bioconductor Biobase (v. 2.0.1) (Gentleman et al., 2004). Quantile normalization and gcRMA background correction were used to derive average expression values for each probe set. Time points were taken every 4 h across the 48-h experiment. Of 37,594 soybean genes represented on the array, 26,006 were called “present” on more than half of the arrays in the experiment using the Affymetrix MAS 5.0 algorithm (also implemented in Bioconductor) and thus were considered to be expressed in developing soybean seeds. To identify genes with a cycling expression pattern, either autocorrelation (Pearson correlation between expression values from 24–48 h and 48–72 h after the initiation of continuous light) or the Haystack tool (Mockler et al., 2007) were used. Models used for Haystack analysis are provided in Supplemental Table 4 and were based on the models from Michael et al. (2008b). Soybean gene models were determined by BLAST search of the Glyma1 version of the soybean genome annotation (Soybean Genome Project, DoE Joint Genome Institute) with the microarray target sequences (http://www.affymetrix.com/ [verified 11 Feb. 2010]). Arabidopsis orthologous sequences represent the top BLASTP match, provided the identity was greater than 30%, from version TAIR8 (ftp://ftp.Arabidopsis.org/home/tair/Genes/TAIR8_genome_release/ [verified 11 Feb. 2010]) of the Arabidopsis genome annotation. It should be noted that soybean is an allopolyploid (Gill et al., 2009) but that the microarray was designed based on expressed sequence tag information and thus while a single Arabidopsis gene will often have a single clear ortholog on the soybean array, this array probe in many cases will hybridize to transcripts from two loci. For the identification of genes with functional annotation related to the soybean transcripts (using Blast2Go [Conesa et al., 2005]), a BLAST e-value cutoff of $10^{-5}$ was used. Microarray data from this experiment is available as series GSE18827 in the Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/ [verified 11 Feb. 2010]). Orthologs of Arabidopsis clock genes in soybean assayed for these experiments were Glyma19g45030 (GmLHY-like), Glyma07g05410 (GmCCA1a), and Glyma06g21120 (GmTOC1a). PCR primers were designed to recognize the soybean transcript with the highest BLASTP identity to the Arabidopsis gene.

Quantitative Polymerase Chain Reaction

For qPCR, 5 μg RNA extracted by the method above was treated with RQ1 DNase (Promega, Madison, WI) as per the manufacturer’s instructions, phenol/chloroform extracted and reverse transcribed with the SuperScript III kit (Invitrogen, Carlsbad, CA) for first strand cDNA synthesis. Samples were diluted 12.5-fold, and 5 μL was used as a template for quantitative PCR with the Stratagene Brilliant 2 SYBR Green kit on a Stratagene MX3000P instrument (Stratagene, La Jolla, CA). Primer sequences are provided in Supplemental Table 4; primers were designed to flank introns in the Glyma1 transcript models and to differentiate a single transcript from the two homeologous loci present in the soybean genome where possible. Relative fold change was calculated by the delta-delta-Ct method using a geometric average of three transcripts for normalization, identified in Supplemental Table 6 as s90, s75, and s84. These control transcripts were chosen because they were amongst the most constantly expressed genes in seeds based on the microarray data. The genes were tested for consistency and amplification efficiency across 15 cDNA samples from the timecourse from both leaf and seed samples, using the GenNorm module (Vandesompele et al., 2002). For data displays of relative gene expression levels in the free-running timecourse, after normalizing the data to control gene levels, the average value at 24 h in continuous light was set to one. For data plots of diurnal timecourse data, expression levels at time zero (lights on) was set to one.

RESULTS

Transcripts in Soybean Seeds are Expressed under Control of the Circadian Clock

To determine what portion of the developing soybean transcriptome is under circadian control, gene expression was profiled in seeds from the early- to mid-filling stage (Goldberg et al., 1989) in free-running conditions. Soybean plants were grown in 12 h light/12 h dark cycles and constant temperature for a total of 10 wk after germination in order to fully entrain the circadian clock. Plants were then introduced to free-running conditions of constant light and temperature for 24 hours before harvesting commenced. Samples were then taken at 4-h intervals over a 48-h time period (Supplemental Fig. 1). Five-week-old whole seeds (embryos and seed coats) were sampled from these plants. To determine whether the soybean orthologs of known clock components oscillate in seed tissues, soybean orthologs of Arabidopsis TOC1, CCA1, and LHY were identified (GmLHY-like, GmCCA1a, GmTOC1a, see methods). Expression of these genes was profiled by qPCR, since there are no probes on the soybean Affymetrix microarray used in this experiment to measure these transcripts. A circadian expression pattern was detected for GmLHY-like and GmCCA1a in both seed and leaf tissue, and with peak expression at or within 4 h of subjective dawn (Supplemental Fig. 2). Although observed oscillations in seeds were of lower amplitude than in leaves, the phase of expression is consistent with the known expression pattern for LHY in Arabidopsis leaf and whole seedlings (Schaffer et al., 1998). Circadian expression for GmTOC1a peaked at circadian time 12 (CT12, 12 h after subjective dawn) in leaf tissue, but robust circadian oscillations of this gene were not detectable in seeds, similar to the pattern previously...
Global transcriptional profiling using the Affymetrix soybean genome array was then performed on the seed RNA samples. To identify which transcripts were circadian clock-regulated, two parallel computational approaches were employed. First, the autocorrelation method was used to identify patterns that recurred between the two days of the timecourse. This method has the advantage that it is insensitive to the waveform of the oscillation (Lin et al., 2002). With this method, 182 probesets with correlation coefficient higher than 0.8 and a fold change value greater than 1.4 were selected for further characterization. Second, average expression values for all genes expressed in seeds were tested using Haystack (Mockler et al., 2007) to identify significant correlation of the timecourse data with a set of canonical rhythmic waveform models (see Methods). Haystack has the advantage of high sensitivity and distinct phase readout, but can only detect oscillations that correlate with one of the canonical models supplied. With Haystack, 481 probesets were identified with a fold change greater than 1.4 and a correlation coefficient higher than 0.8 with one of a number of different circadian expression models of defined phase at 1-h resolution. The intersection of these two sets, containing 131 probesets, was identified as circadian-regulated by both methods. The EST-derived exemplar sequence to which each probeset on the array was originally designed was compared to the current soybean gene models (Glyma1 release, DoE Joint Genome Institute, www.phytozome.org [verified 11 Feb. 2010]). A total of 42 gene models from the current soybean genome matched more than one probeset (indicating that duplicate probesets exist for the transcript on the Affymetrix array). Duplicate probesets were then removed, and only one probeset was considered in further analysis (to avoid over-representation of these genes in promoter element enrichment statistics, or in the data displays for multiple transcripts). The de-replicated lists of genes identified as cycling by Haystack together with the additional genes identified by autocorrelation (a total of 479 genes) are given in Supplemental Table 1. Since the Arabidopsis genome is the best annotated whole-genome sequence of a dicotyledon plant, predicted soybean proteins were compared to Arabidopsis proteins using BLASTP, and the top Arabidopsis protein hits are also listed in Supplemental Table 1. All but 11 of the soybean sequences identified shared more than 30% predicted amino acid level identity with a likely ortholog in the Arabidopsis protein set.

Transcripts Encoding Proteins with Distinct Roles in Seed Metabolism and Biochemistry are Segregated by Phase

Circadian-regulated transcripts were assigned to bins for phase of peak expression either automatically by correlation to a circadian pattern (by Haystack) or manually in the case of the genes identified by autocorrelation. A clear phase of peak expression could be identified for all but 11 of the genes. The genes were not randomly distributed with respect to phase; most genes fell into one of two major phase bins (Fig. 1) with the largest number of genes falling in the bin peaking at CT20 and a smaller number of genes peaking at CT4-CT8. To identify pathways that may be influenced by circadian regulation in seeds, MAPMAN was used to visualize the expression of the Arabidopsis orthologs of circadian clock-regulated soybean genes (Thimm et al., 2004). While not all of the soybean cycling genes could be placed into a MAPMAN pathway in this manner, genes involved in light harvesting and the Calvin cycle were found to be co-regulated, with a peak in expression at CT20, and a minimum expression level at CT8 (Fig. 2). To obtain additional functional characterization for the circadian clock-regulated genes, Blast2Go was used to obtain approximate homology-based gene functions, which makes possible the association of soybean gene models with gene ontology (GO) terms from annotation in other species (Conesa et al., 2005). For 257 of the 479 genes one or more GO slim Biological Process terms were identified, which are listed in Supplemental Table 2. From this analysis, 14 genes with predicted functions in lipid metabolism were identified in the set of circadian clock-controlled genes, with peak expression for most of these genes occurring after 36 h in continuous light (at CT12-CT24, Fig. 3A). There were two classes of genes predicted to be involved in carbohydrate metabolism. Of the carbohydrate metabolism genes, the subset which are also predicted to be involved in photosynthesis show co-ordinate regulation with expression peaking at 40 to 44 h after transfer to continuous light (Fig. 3B, peak at CT16-CT20). The remainder of the genes predicted to function in carbohydrate metabolism, but without GO annotations associated with photosynthesis, peak at various times during the circadian cycle (Fig. 3C).
A Subset of Circadian-regulated Transcripts in Developing Soybean Seed Are Phase Shifted with Respect to Leaf Transcripts

The peak expression for the photosynthetic and light-harvesting transcripts in seeds was observed to be CT19-21 h, or close to subjective midnight, a result that is unexpected for transcripts involved in producing components of the photosynthetic machinery (see Fig. 2, Supplemental Fig. 3). In previous reports of circadian gene expression in Arabidopsis seedlings (Harmer et al., 2000), these transcripts were shown to peak at CT4-5. In order to investigate this discrepancy, seed and leaf gene expression of photosynthetic transcripts were compared in parallel from the same soybean plants grown in free-running conditions by qPCR (Fig. 4A and B, Supplemental Table 6). In soybean leaves, genes annotated as encoding components of photosystem I (PSI) or photosystem II (PSII) peak just before subjective noon in the phase CT4-5, which is similar to the peak phase observed for transcripts encoding PSI and PSII components in whole Arabidopsis seedlings, and expected of genes involved in producing proteins involved in light harvesting.

Diurnal Regulation of Photosynthetic Transcripts in Soybean Seeds

In order to determine whether the near-antiphasic expression of photosynthetic transcripts in seeds with respect to leaves was restricted to free-running conditions, the expression of the PSI and PSII transcripts was examined in seeds and leaves maintained in diurnal conditions (the 12 h light/dark cycles and constant temperature used to entrain the circadian experiment). While the overall amplitude of the oscillations of PSI and PSII transcripts are smaller in seeds than in leaves, the transcripts clearly oscillate in both seeds and leaves in these conditions and are nearly in phase with one another (Fig. 4C-F). This indicates that in diurnal conditions seed and leaf circadian gene expression is synchronized, and the peak phase of photosynthetic gene expression occurs just before subjective noon.

Circadian and Diurnal Regulation of Genes Involved in Carbohydrate Metabolism

To validate the microarray expression profiles for genes involved in carbohydrate metabolism, several genes, including PHOSPOGLUCOMUTASE (PGMP) and CELLULOSE SYNTHASE A (CESA) were profiled in circadian and diurnal conditions in seeds and leaves using...
qPCR. Figure 5A and B show that expression of PGMP and CESA peaks in seeds at CT4-8 (28–32 h after transfer to continuous light) and in leaves at CT24 (Supplemental Table 6). For both of these transcripts peak expression is observed 4 h after dawn in diurnal conditions. For GENOMES UNCOUPLED4, a magnesium chelatase cofactor involved in chlorophyll biosynthesis and the coordination of plastid and nuclear gene expression (Larkin et al., 2003), and a gene encoding a subunit of GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE A, qPCR on seed and leaf tissues validated the circadian expression pattern observed in the microarray.

Fig. 3. Circadian co-regulation of genes by biological process. Normalized microarray expression profiles of soybean genes grouped by predicted functional category [determined by Gene Ontology [GO] terms]. (A) Transcripts predicted to encode proteins involved in lipid metabolism, (B) transcripts predicted to be involved in carbohydrate metabolism reactions of photosynthesis, or (C) transcripts encoding genes predicted to be involved in carbohydrate metabolism reactions that are not directly related to photosynthesis. For all genes, the highest average value is set to 1. Abbreviations: OPR2: 12-OXOPHYTODIENOATE REDUCTASE 2, SQE3: SQUALENE EPOXIDASE 3, LOX1: LIPoxyGENASE1, GID1C: GA INSENSITIVE DWARF1C, LPP2: LIPID PHOSPHATE PHOSPHATASE 2, GAMMAVPE: GAMMA VACUOLAR PROCESSING ENZYME, OPR1: 12-OXOPHYTODIENOATE REDUCTASE 1, FAD2: FATTY ACID DESATURASE2, SBPASE: SEDOHEPTULOSE-BISPHOSPHATASE, RBCS: RIBULOSE BISPHOSPHATE CARBOXYLASE SMALL, GAPA-1: GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE A, PRK1: PROTEIN PHOSPHATASE 7, PGMP: PHOSPHOGLUCOMUTASE, APL3: ADPGLUCOSE PYROPHORYLASE LARGE SUBUNIT, SBE2.2: STARCH BRANCHING ENZYME 2.2.
experiment and suggest that the timing of peak expression is not synchronized in seeds and leaves in free-running conditions (Supplemental Fig. 4).

Promoter Elements in Circadian-Regulated Genes

The coordinated expression pattern of seed circadian transcripts led to an investigation of whether the distinct phases of gene expression are controlled by particular cis-regulatory elements in the promoters of circadian clock-controlled genes in seeds. In order to identify promoter motifs that were over-represented within groups of soybean seed genes expressed in different phases, the Elefinder module (http://stan.cropsci.uiuc.edu/cgi-bin/elefinder/soyelefinder.cgi [verified 11 Feb. 2010]) (Pan et
al., 2009) was used to compare the abundance of known plant regulatory motifs in sets of genes in each 1-h phase bin with the abundance of these motifs in the promoters of all genes measured by the microarray, and rank motifs on the basis of over-representation (using an uncorrected \( P \) value generated by the binomial distribution) in the genes showing co-ordinated expression (Hudson and Quail, 2003; Pan et al., 2009). Since an uncorrected \( P \) value is generated for each element examined, and approximately 100 different known elements are currently in the database used by the software, a stringent multiple testing correction (the Bonferroni correction) suggests that all \( P \) values below \( 5 \times 10^{-4} \) should be treated as significant at the 0.05 level. The most significantly over-represented elements in the clock-controlled genes from soybean seeds include known light-responsive and circadian motifs such as the GATA, CCA1 binding site, and the SORLIP1 motifs (Giuliano et al., 1988; Wang et al., 1997; Harmer et al., 2000; Hudson and Quail, 2003). The GATA element was significantly enriched in the promoters of circadian genes expressed in several different phase windows, but especially in genes peaking in expression at CT20, with a significance value of \( 1.7 \times 10^{-20} \), and the SORLIP1 occurred preferentially in promoters of genes peaking in expression between CT19 and CT2 (Fig. 6). The GATA motif has previously been shown to be enriched in the promoters of Arabidopsis seedling and leaf genes that peak late in the afternoon (Covington et al., 2008).

Another known motif over-represented in this set of promoters was the abscisic acid responsive element (ABRE, \([T/C]ACGTG(GC)\) and ABRE-like \([IC/G/T\]ACGT(G/T)\) motifs, which are a class of G-box-like elements implicated in responses to abscisic acid and known to be recognized by bZIP proteins (Marcotte et al., 1989; Guiltinan et al., 1990). Two soybean transcripts homologous to the Arabidopsis HB-zip transcription factors ATHB2 and ATHB6 (see Supplemental Table 1) (Soderman et al., 1994) were found to be circadian clock-regulated in seeds (peaking at CT4 or CT5). The known binding site for ATHB2 is, intriguingly, somewhat over-represented in genes peaking at CT6 and CT11 (Fig. 6).

The CCA1-binding site (\([A\{A/C\]}AATCT\) (Wang et al., 1997) was found to be overrepresented in promoters of genes that peak at CT01 (near subjective dawn) and in genes that peak at CT11 (near subjective dusk) in this experiment (Fig. 6). At subjective dusk the genes containing the CCA1-binding site in their promoters include the orthologs of SOS1, TGA4/OBF4, and two additional genes with annotations as water stress-induced transcripts—all genes with putative functions in cold and desiccation tolerance, consistent with the role of the circadian clock in regulating resistance to lower temperatures at night (Buttner and Singh, 1997; Shi et al., 2000; Chen et al., 2002) (Supplemental Table 3). In the subjective dawn-phased genes, the CCA1 binding site is found in the promoters of \textit{PATTY ACID DESATURASE 2} (PAD2) and other genes involved in biosynthesis and metabolism (Okuley et al., 1994). The CCA1 binding site and the morning element (AACCAC; Harmer and Kay, 2005) were also represented in the promoters of photosynthesis-related genes peaking at CT20 in seeds, however these were not found to be as strongly over-represented as the motifs discussed above. The over-representation of the CCA1 binding site was found to be statistically significant with the Bonferroni correction, but the over-representation of the evening and morning elements were not found to be significant in this experiment.

Datasets from \textit{Arabidopsis} circadian microarray experiments (from either leaves or seedlings) were interrogated from the DIURNAL database (Mockler et al., 2007) (http://diurnal.cgrb.oregonstate.edu/ [verified 11 Feb. 2010]) in order to compare \textit{Arabidopsis} gene expression to that of the predicted orthologous cycling genes in soybean seeds. For 221 soybean genes, the \textit{Arabidopsis} ortholog showed a circadian expression pattern in leaves or seedlings. For another 236 soybean genes, the closest \textit{Arabidopsis} gene did not show a circadian expression pattern, and 22 were not represented in the dataset (either no significant sequence match for the gene was found in \textit{Arabidopsis} or the \textit{Arabidopsis} orthologs for these genes were not interrogated by the \textit{Arabidopsis} microarrays) (Mockler et al., 2007). Thus, only about half of the genes that were found to cycle in soybean seeds in this experiment would be predicted as circadian based on the expression pattern of their orthologs in \textit{Arabidopsis} seedlings and leaves. For further functional insight into the cycling genes expressed in soybean seeds, the expression patterns of the \textit{Arabidopsis} orthologs were visualized using the Bio-Array Resource to determine if these genes are preferentially expressed in developing seeds (Supplemental Fig. 5) (Toufighi et al., 2005).

In general, expression of the genes that cycled in both soybean seeds and \textit{Arabidopsis} seedlings was higher in leaf tissues than in seeds. Some of the genes that did not cycle in \textit{Arabidopsis} whole seedlings (but show a circadian pattern of expression in soybean seeds) appear to be enriched in \textit{Arabidopsis} seed with respect to leaf tissues (Supplemental Fig. 5B, for example At5g44120 [\textit{CRUCIF-ERINA}] or At5g03860). While many of the soybean seed cycling genes do not appear to be specific to seed tissues in \textit{Arabidopsis}, those that are may represent a conserved, seed-specific group of circadian-expressed genes.

**DISCUSSION**

It has been shown that 6 to 40% of the transcriptome is under circadian control in \textit{Arabidopsis} seedling or leaf tissues, and up to 90% of genes in \textit{Arabidopsis} have been shown to cycle in some circadian or diurnal conditions (Harmer et al., 2000; Edwards et al., 2006; Covington et al., 2008; Michael et al., 2008b). In contrast, seed transcripts in soybean with a statistically significant recognizable circadian expression pattern using similar criteria make up only 1.8% of the seed transcriptome, which is similar to the observation that fewer genes cycle in \textit{Arabidopsis} roots than in shoots (3% in roots...
vs. 14% in shoots in James et al., 2008). In addition, the amplitude of oscillation in seeds is lower than in leaves in both diurnal and circadian conditions. However, circadian variations in expression are clear for a number of transcripts that are predicted to be involved in several biological processes crucial for seed development and composition, including lipid metabolism and photosynthesis. The preponderance of genes controlled by the circadian clock in soybean seeds are expressed in one of two distinct phase intervals: CT19-21 or CT4-8 (Fig. 1), which is similar to the pattern of circadian clock–controlled gene expression in whole Arabidopsis seedlings (Blasing et al., 2005; Harmer et al., 2000; Michael et al., 2008b). The identification of conserved, over-represented promoter motifs in these sets of genes may prove useful for further dissection of the regulatory mechanisms underlying seed gene expression.

A number of genes in seeds are expressed under circadian control in free-running conditions, which implies circadian regulation of the biochemical processes such as light capture, carbon fixation, and influence over fatty acid synthesis, and seed development. Consistent with this observation, a number of regulatory motifs previously correlated with circadian clock–controlled gene expression have been identified in the promoters of these genes. Finally, oscillations are detectable at the mRNA level for the core clock transcription factors CCA1 and LHY, although not for TOC1, which is consistent with the presence of an active but incomplete oscillator in seeds. The limited set of genes observed to oscillate in expression in soybean seeds are potentially downstream of oscillations in the clock “morning” genes such as LHY and CCA1. Some evidence for this hypothesis is given by the observation that the CCA1 binding site is over-represented in the genes which show cycling expression patterns in seeds.

Approximately half the genes with circadian-regulated expression patterns in soybean seed have circadian-regulated orthologs in whole Arabidopsis seedlings. For the remainder of the genes, the Arabidopsis orthologs of circadian genes from soybean seeds do not show a strong circadian expression pattern in whole seedlings or leaves. In some cases, based on analysis of available microarray expression data from Arabidopsis, the soybean seed genes that do not cycle in Arabidopsis are expressed specifically at higher levels in Arabidopsis seed tissues. Therefore oscillations in these genes may not be detectable in circadian array experiments performed on leaf tissues or whole seedlings. Another possibility which cannot be ruled out from these data is that while the soybean gene measured in this experiment cycles, the closest ortholog in Arabidopsis is not clock-regulated. A final possibility is that, due to imperfect annotation of the soybean genome, the orthology has been incorrectly assigned.

In free-running, continuous light conditions, transcripts for some genes including PSI and PSII components as well as genes involved in carbohydrate metabolism appear to peak in expression at different times in seeds and leaves from the same plants. However, in diurnal conditions, which approximate the normal field conditions for developing soybeans, the photosynthetic transcripts are expressed in the same, expected phase in both leaves and seeds, and thus the change in phase appears to be a result of the shift to continuous light conditions and is confined to seeds. The plant circadian clock is capable of organ autonomous phasing and periodicity (Thain et al., 2000, 2002). The underlying cause for the observed shift in phase in seeds under free-running conditions is not known; however, the observation is consistent with several other studies. In developing maize kernels, the Opaque2 transcript undergoes oscillations in expression that persist.
when the ear is covered in foil, suggesting that direct light perception by the seed is not required to maintain this expression pattern and implicating the involvement of signals from the maternal plant (Cicieri et al., 1999). In Arabidopsis roots, differences in the circadian clock–controlled expression pattern of genes between free-running, continuous light conditions and diurnal conditions are exacerbated by providing the plants with sucrose (James et al., 2008). Many genes, including circadian and light-regulated genes, are strongly influenced by diurnal changes in sugar availability (Blasing et al., 2005; Usadel et al., 2008). In sink tissues such as seeds and roots, it is possible that expression of the clock-controlled photosynthesis-related genes may be especially responsive to carbohydrate levels, and the regulation of these genes by the circadian clock and other factors will require further investigation. Another possibility is that specific changes in the metabolic state of the seed arising from the switch to continuous light conditions and diurnal conditions are sufficient to affect the phasing or periodicity of clock-controlled gene expression in seeds.

Here it is shown that genes likely to have a key role in seed composition are regulated in both circadian and diurnal cycles in developing soybean seeds. The effect of these factors under natural conditions, where day lengths change incrementally throughout the duration of the growing period, remain to be characterized. The circadian effects on the machinery of seed filling, and the role of a dark period, remain to be characterized. The circadian effects on seed filling metabolism, and the role of a dark period, are thus physiological parameters that may play a role in seed development hitherto not widely recognized.

Acknowledgments

The author would like to thank Dr. Matthew Hudson and Dr. Larry Dunkle for critical reading of the manuscript. Product names are necessary to report factually on available data. However, the USDA neither guarantees nor warrants the standard of the product, and the use of the names implies no approval of the product to the exclusion of others that may also be suitable. Funding for this work was provided through the USDA-ARS Current Research Information System 3602-21000-004-00D.

References


