Integrating ELF4 into the circadian system through combined structural and functional studies

Integrating *ELF4* into the circadian system through combined structural and functional studies

Elsebeth Kolmos,[1,8] Monika Nowak,[2] Maria Werner,[3] Katrin Fischer,[4] Guenter Schwarz,[4] Sarah Mathews,[5] Heiko Schoof,[1] Ferenc Nagy,[6,7] Janusz M. Bujnicki,[2,3] and Seth J. Davis[1]

[1]Max Planck Institute for Plant Breeding Research, Carl-von-Linné-Weg 10, 50829 Cologne, Germany
[2]Faculty of Biology, Institute of Molecular Biology and Biotechnology, Adam Mickiewicz University, Umultowska 89, 61-614 Poznan, Poland
[3]Laboratory of Bioinformatics and Protein Engineering, International Institute of Molecular and Cell Biology, Trojdena 4, 02–109, Warsaw, Poland
[4]Institute of Biochemistry, University of Cologne, Otto-Fischer-Strasse 12-14, 50674 Cologne, Germany
[5]Arnold Arboretum of Harvard University, 22 Divinity Avenue, Cambridge, Massachusetts 02138
[6]Institute of Plant Biology, Biological Research Centre of the Hungarian Academy of Sciences, 6726 Szeged, Hungary
[7]School of Biological Sciences, Edinburgh University, Mayfield Road, Edinburgh EH9 3JH, United Kingdom
[8]Present address: Section of Cell and Developmental Biology, Division of Biological Sciences, University of California San Diego, La Jolla, California 92093

(Received 24 April 2009; accepted 7 August 2009; corrected 23 October 2009; published online 22 October 2009)

The circadian clock is a timekeeping mechanism that enables anticipation of daily environmental changes. In the plant *Arabidopsis thaliana*, the circadian system is a multiloop series of interlocked transcription-translation feedbacks. Several genes have been arranged in these oscillation loops, but the position of the core-clock gene *ELF4* in this network was previously undetermined. ELF4 lacks sequence similarity to known domains, and functional homologs have not yet been identified. Here we show that ELF4 is functionally conserved within a subclade of related sequences, and forms an alpha-helical homodimer with a likely electrostatic interface that could be structurally modeled. We support this hypothesis by expression analysis of new *elf4* hypomorphic alleles. These weak mutants were found to have expression level phenotypes of both morning and evening clock genes, implicating multiple entry points of ELF4 within the multiloop network. This could be mathematically modeled. Furthermore, morning-expression defects were particular to some *elf4* alleles, suggesting predominant ELF4 action just preceding dawn. We provide a new hypothesis about ELF4 in the oscillator—it acts as a homodimer to integrate two arms of the circadian clock. [DOI: 10.2976/1.3218766]

CORRESPONDENCE
Seth J. Davis:
davis@mpiz-koeln.mpg.de

Many organisms possess a biological clock that facilitates anticipation to daily environmental changes. The chronobiological mechanism that drives this anticipation is termed the circadian clock, and it generates rhythms that have a duration of about 24 h. The environmental light-dark (LD) cycle is the major factor that directs the synchronization of daily biological rhythms. However, in many geographical locations, the duration of day-length changes over the year as a consequence of the tilted axis of the Earth. This prompts the circadian clock to be reset every day, to be in sync with daylight changes over the seasons, and this process is termed entrainment. In plants, the circadian clock is necessary to synchronize metabolic and developmental processes (Dodd *et al.*, 2005; Fukushima *et al.*, 2009; Salome *et al.*, 2008). Much progress has been made to understand the organization of the transcription-

translation feedback loops underlying this mechanism in the model plant *Arabidopsis thaliana* (Arabidopsis). Progress on entrainment of the oscillator is also emerging (Harmer, 2009). The long-term goal is to fully account for the genetic factors that bridge the transitions of light-perception, which occur at dawn and dusk, and to integrate these features into the oscillator mechanism.

In a recent model of the plant-circadian system, central-clock genes were arranged in three interconnected feedback loops (Locke *et al.*, 2006; Zeilinger *et al.*, 2006). In the core loop, the Myb-like morning transcription factors *CIRCADIAN CLOCK ASSOCIATED 1* (*CCA1*) and *LATE ELONGATED HYPOCOTYL* (*LHY*) repress transcription of the pseudoresponse regulator (PRR) *TIMING OF CAB EXPRESSION 1* (*TOC1*) (Alabadi *et al.*, 2002; Perales and Mas, 2007; Schaffer *et al.*, 1998; Strayer *et al.*, 2000; Wang and Tobin, 1998). Specifically, this morning repression results in a restriction of *TOC1* expression, and only as repression is relieved does evening activation occur. Closing the core of the clock, *TOC1* is believed to confer a genetic activity that ultimately results in the promotion of *CCA1* and *LHY* transcription at the end of the night (Alabadi *et al.*, 2001, 2002). It was recently established that the TCP transcription factor *CCA1 HIKING EXPEDITION (CHE)* links the feedback activity of *TOC1* to the transcriptional induction of *CCA1* (Pruneda-Paz *et al.*, 2009). A morning loop, that includes *PRR9* and *PRR7*, is interlocked with the core loop. *CCA1*/*LHY* promotes *PRR9* and *PRR7* expression, which in return have negative feedback on *CCA1*/*LHY* (Farre *et al.*, 2005; Nakamichi *et al.*, 2005a, 2005b). Separately, an evening loop containing *GIGANTEA (GI)* (Fowler *et al.*, 1999) as a partial constituent of a mathematically defined "*Y*" component, reciprocally connects to *TOC1* expression. The evening-expressed *LUX ARRHYTHMO* (*LUX*) encodes a Myb-containing transcription factor that likely functions in a feedback role similar to *TOC1* (Hazen *et al.*, 2005; Onai and Ishiura, 2005), and it remains a possibility that *LUX* could also be a component of the *Y* activity. In addition, an unknown gene *X* mathematically is predicted to bridge *TOC1* activity in the promotion of *CCA1*/*LHY* (Locke *et al.*, 2006; Zeilinger *et al.*, 2006).

Additional genes must be considered in the oscillator mechanism. For example, we have previously hypothesized that dusk expression of *EARLY FLOWERING 4* (*ELF4*) is interlocked with the central *CCA1/LHY-TOC1* loop (Kolmos and Davis, 2007; McWatters *et al.*, 2007). *ELF4* is necessary and sufficient to promote both *CCA1* and *LHY* and repress *TOC1* (McWatters *et al.*, 2007). Importantly, the modularity of the multiloop structure of the circadian system is believed to enable robust and dynamic gene regulation, which is stable under divergent environmental conditions (Herzog, 2007). Positioning genes genetically known to function within the oscillator is thus a critical task.

In the absence of *ELF4*, the circadian clock fails to sustain rhythmicity under constant conditions and the residual clock rhythms are imprecise (Doyle *et al.*, 2002; Kikis *et al.*, 2005; McWatters *et al.*, 2007). This results in the plants being early flowering, and the *elf4* lines have an elongated hypocotyl that is particularly evident after growth under a short-day photoperiod. Both of these phenotypes have been hypothesized to be due to a defect in daylength-sensing (Dowson-Day and Millar, 1999; Doyle *et al.*, 2002; Nozue *et al.*, 2007). Earlier investigations of *ELF4* were based on *elf4* loss-of-function mutants (Doyle *et al.*, 2002; Khanna *et al.*, 2003; Kikis *et al.*, 2005; McWatters *et al.*, 2007). The first allele was isolated in the Wassilewskija-2 accession (Ws-2, Ws hereafter), termed *elf4-1*, and it has a deletion of the *ELF4* coding region (Doyle *et al.*, 2002). The two Columbia-0 (Col-0) *elf4* alleles, *elf4-101* and *elf4-102*, are T-DNA insertion lines that display a null phenotype that is phenotypically less severe compared to *elf4-1*, regarding amplitude of *CCA1* transcript accumulation and hypocotyl length (Khanna *et al.*, 2003; Kikis *et al.*, 2005). In these pioneering studies on *elf4*, the *ELF4*-encoded polypeptide remained effectively uncharacterized.

In Arabidopsis, *ELF4* belongs to a small gene family and has four sequence homologs (*EFL1–EFL4*). These sequences lack detectable amino-acid similarity to characterized proteins outside the ELF4 superfamily and, more generally, this family has not been detected outside the plant kingdom (Boxall *et al.*, 2005; Doyle *et al.*, 2002; Khanna *et al.*, 2003). For this reason, sequence repositories have classified the ELF4 family as its own functional class; InterPro DUF1313 (http://www.ebi.ac.uk/interpro/IEntry?ac=IPR009741). A limited comparison of ELF4-like sequences revealed several conserved residues in the ELF4 family (Doyle *et al.*, 2002; Khanna *et al.*, 2003). About 20% of the ELF4 amino-acid residues were shown to be conserved, and two subgroups could be defined. One group contained the Arabidopsis sequences ELF4 and EFL1, and the second group included EFL2, EFL3, and EFL4. Sequences from rice and sorghum were on a branch with EFL2, EFL3, and EFL4 (Khanna *et al.*, 2003), suggesting that a rice ortholog of ELF4 remained to be detected. In contrast, analyses that included McELF4 from iceplant showed that this sequence is on a branch with ELF4, identifying it as an ortholog (Boxall *et al.*, 2005). Transcriptional expression data are consistent with phylogenetic predictions. *McELF4* is clock-controlled at the level of transcript accumulation, whereas the rice *ELF4*-like genes apparently are not (Boxall *et al.*, 2005; Murakami *et al.*, 2007). Thus, the functional conservation of ELF4 is unclear and should be clarified.

In this study, we placed *ELF4* within the circadian network using molecular-genetic approaches. We complemented this by informatic approaches to aid in defining ELF4 functionality. Together, these methods enabled us to measure ELF4 conservation and propose the DUF1313 functional domain. Notably, a proposed homodimer and α-helical

structure of ELF4 was empirically confirmed. We then characterized new hypomorphic *elf4* alleles that we identified. For this, we took advantage of a targeted screen for Ethyl methanesulfonate (EMS)-mutations (Targeting Induced Local Lesions IN Genomes, TILLING) in *ELF4* to generate an allelic series of mutants. In this way, we could systematically dissect the signal network of ELF4 action within the oscillator system. In the *elf4* TILLING analysis, we isolated three strong reduced-function alleles that corresponded to residues predicted to be on a surface-charged structure. Additionally, we found that the expression of clock genes was altered throughout the *elf4* TILLING collection. Importantly, the subtle *elf4* alleles helped the conclusion that *ELF4* is important for down-regulation of *PRR7* expression during the night. Using this information, we could place *ELF4* within the oscillator mechanism, which could be replicated mathematically. From our detailed structural-functional studies, we propose that ELF4 functions as a small dimer critical for buffering perturbations to the evening-arm of the circadian clock. Generally, our work provides a comprehensive approach to the integration of proteins of unknown function within a structural and functional framework of their signaling pathway.

RESULTS

ELF4 is a predicted one-domain protein

The structure of ELF4 is previously undetermined. Accordingly, we used a phylogenetic approach to propose the ELF4 functional domain. ELF4-related sequences from 25 species were identified by basic local alignment search tool (BLAST) in expressed-sequence tagged (EST) cDNA collections and in genomic databases; no ELF4-like sequences were detected outside the plant kingdom. These clones were obtained from various genome projects, and full-length sequences were determined. We used these to generate a multiple alignment (see Supplementary Material, Fig. S1). This revealed a central conserved region between the ELF4/EFL-related polypeptides comprising ELF4-residues 21–93 (the DUF1313 domain), of which a short stretch (residues 58–63) was less conserved. Furthermore, two subgroups could be identified within the alignment: one encompassed an ELF4-related group and the other an EFL-related group. We noted that all predicted polypeptides were ~110–150 amino acids in length, and all were individual domain proteins. Also, outside the central core of conservation, all ELF4-related polypeptides had both amino- and carboxy-terminal extensions. These "tails" were more divergent between the ELF4 and the EFL subgroups. As all ELF4 sequences were most similar within the central part of the protein, this could be the critical domain for structure and function.

There was a high degree of sequence conservation within the ELF4 family (see Supplementary Material, Fig. S1). The ELF4 group was divergent from the EFLs primarily in the most N- and C-terminal regions, and within the ELF4 group,

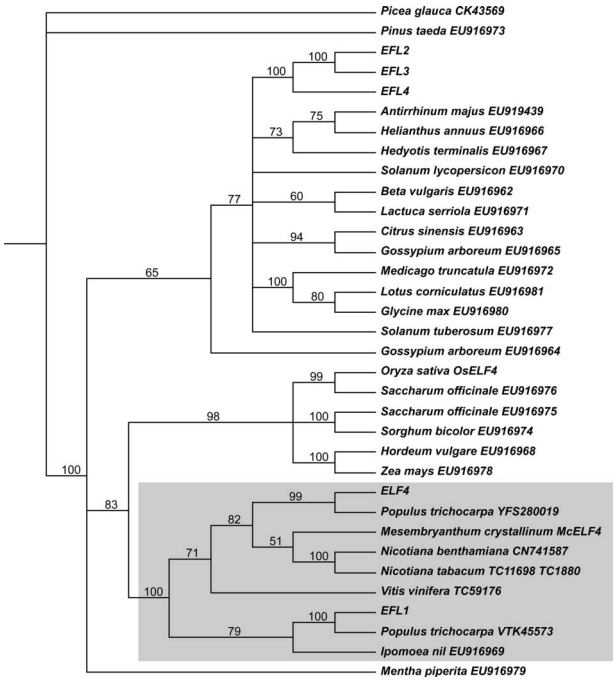

Figure 1. *ELF4* phylogeny. *ELF4* Bayesian consensus tree. Majority-rule consensus of 3601 trees (burn-in=402 trees) from two runs of 2 million generations each. Bayesian posterior probabilities indicated support for individual nodes are above the branches as percentages. The *ELF4/EFL1* clade is shaded in gray.

the previously hypothesized ELF4 nuclear localization site (KRRR; residues 8–11) was not conserved (Khanna *et al.*, 2003). Fifteen residues in the central part of ELF4 were fully conserved across the whole family (see Supplementary Material, Fig. S1), a refinement compared to the earlier study (Khanna *et al.*, 2003). Some acidic amino acids (S38, Q48, Q49, and Q55) characterized the ELF4 subgroup followed by a conserved glycine (G74), a serine-lysine pair (sites 77–78), and another serine residue (S84). The conserved residues in the ELF4 family are likely to be important for structural stability and/or for function.

We performed phylogenetic analyses in order to determine the orthology of *ELF4* and related sequences in Arabidopsis and other angiosperms. The consensus tree from the Bayesian analysis (Fig. 1) highlights clades that were supported by the data in all the analyses. Notably, one of the well-supported clades contains both *ELF4* and *EFL1*, each of these sequences occurring in a separate clade along with their respective homologs from other eudicot sequences. The other *EFL* sequences (*EFL2–EFL4*) represent a recent diversification in a more distantly related gene lineage (Fig. 1). The other clades that receive high support consist of sequences sampled from a single family (grasses, legumes, and tobaccos). There are many sequences from single families or relatively closely related groups that do not form clades in any of the trees. For example, the *EFL* sequences from the two *Solanum* and the two *Gossypium* species are not closely

related, nor are those from Asteraceae (*Helianthus* and *Lactuca*), nor from asterids (*Solanum*, *Helianthus*, *Lactuca*, *Hedyotis*, and *Antirrhinum*) and from rosids (Arabidopsis, *Citrus*, *Gossypium*, and the legumes). This suggests a complex gene history of duplication and loss. The position of the clade of grass sequences as sister to the *ELF4*/*EFL1* clade is moderately supported in the Bayesian (posterior probability of 0.83) and parsimony (82% bootstrap value; data not shown) trees, and is not supported in the likelihood bootstrap consensus tree (data not shown). As in eudicots, there is evidence of gene duplication in the grasses with, for example, multiple sequences occurring in sugar cane (Fig. 1) and rice (Murakami *et al.*, 2007). Thus, the sequences from grasses appear to be related to the eudicot *ELF4*/*EFL1* lineage, but independent gene diversification in grasses makes it difficult to predict which sequence or sequences might encode ELF4-like function.

ELF4 and EFL structure prediction, ELF4 dimer, and structure refinement: we sought to expand on the structural analysis of the ELF4/EFL family through computational analysis (see Supplementary Material, Fig. S2A). For this we started by inferring the protein folds of the DUF1313 sequences we described above through the direct use of *de novo* protein structure prediction method ROSETTA. This automated modeling was conceptually similar to the method we used previously (Kolmos *et al.*, 2008). We obtained a consensus prediction from the analysis of all five Arabidopsis and an additional 20 non-Arabidopsis ELF4/EFL proteins. All these proteins were predicted to be α-helical with disordered N- and C-termini, and the sequence-conserved core folds back into a zipper-like confirmation. The nonconserved tails were all predicted to be soluble and to fluctuate between many alternative conformations. Note that the regions of the α-helical fold are exactly the sequence interval conserved in this family (see Fig. 1 and Supplementary Material, Fig. S2A). That ROSETTA consistently generated similar structures was taken as a strong predictor that our modeling represents the energetically stable native protein fold of ELF4/EFL. We thus decided to obtain a more statistically accurate model focusing on ELF4 (from Arabidopsis) through an expanded and integrated use of computational and biochemical methods.

We performed advanced protein structure prediction on ELF4 using the GeneSilico metaserver. All methods tested led to a region, comprising residues 21–93, as forming two α-helices with termini that were intrinsically disordered. This was exactly what we predicted above. This result was further supported by several methods for coiled-coil prediction (see Methods). Finally, *de novo* structure prediction, with a suite of algorithms, indicated that the central region of ELF4 preferentially folds into a coiled-coil structure (see Supplementary Material, Fig. S2B). Here, as above, the residues 21–57 fold into a helix, the small region 58–63 forms a loop, and 64–93 folds into a second helix (see Supplementary Material, Figs. S1 and S2B). These helices interact with each other. Additionally, the amino- and carboxy-terminal tails extend away from the folded core. We conclude that this is the general monomer fold for all DUF1313 proteins.

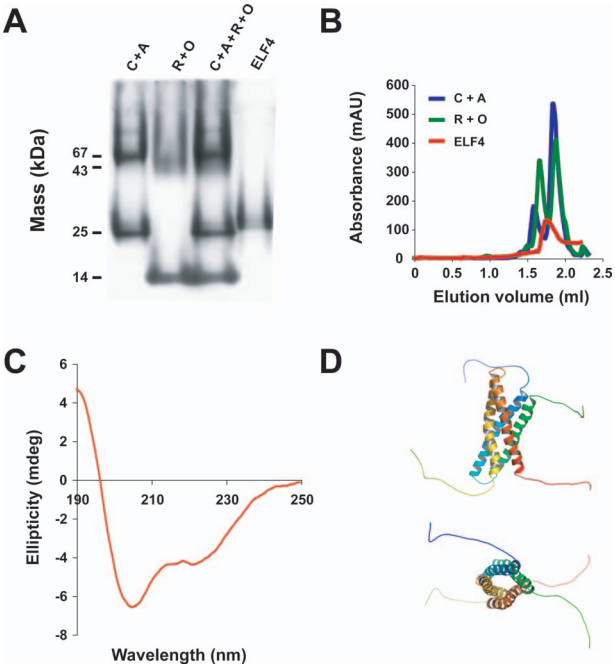

Figure 2. ELF4 self-associates to form a homodimer. The molecular mass of the ELF4 was empirically determined. The ELF4 dimer was subsequently predicted. (A) Purified ELF4 protein migrates with a size of ~26 kDa after blue native-polyacrylamide gel electrophoresis (BN-PAGE). The bands were visualized after Coomassie staining. Standards: A, albumin (67 kDa); C, chymotrypsinogen A (25 kDa); O, ovoalbumin (43 kDa); R, ribonuclease A (13.7 kDa). (B) Gel filtration of purified recombinant ELF4. Standards as in (A). Note that ELF4 elutes exclusively as a single peak between the 13.7 kDa and 43 kDa standards, indicating a tight homodimer. (C) CD spectrum of purified ELF4 confirms that the native fold of ELF4 is mainly α-helical, as indicated by the two negative peaks around 222 nm and 205 nm. (D) ELF4 dimer model. ELF4 self-associates with a dimer interface along the α-helical fold. Side view (top) and end-on view (bottom) of the proposed ELF4 dimer. The two ELF4 monomers are colored yellow-red and blue-green, respectively.

To further refine our model, we needed to establish if there was a docking interface in ELF4. To test this, we assessed biochemically if ELF4 could self-associate. First we purified recombinant ELF4 protein and analyzed its migration speed by native polyacrylamide gel electrophoresis (PAGE). As can be seen in Fig. 2(A), ELF4 migrates at ~26 kDa, which is the size of a homodimer. To corroborate the ELF4 homodimer, we assessed the flow-rate of ELF4 on a sizing column. Here, we also found ELF4 to be a tight homodimer [Fig. 2(B)].

To determine if the secondary structure of native ELF4 conformed to the structural predictions, we obtained far-ultraviolet (UV) CD spectra. The averaged spectra revealed a

strong α-helical signal and disordered signals, and no contributions of β-strands were evident [Fig. 2(C)]. ELF4 was thus found to be in a mixture of helical and disordered folding states.

In accordance with our biochemical findings that ELF4 is an α-helical dimer, we sought to define computationally a probable dimer interface. We performed independent docking analyses for the core region (residues 22–95) with the algorithms GRAMM, DOT, and ZDOCK. The consensus result revealed a symmetrical antiparallel orientation of two monomers. This provided a platform for further model refinement *via* optimization with ROSETTADOCK. According to the PROQ model quality assessment method, the dimer model (comprising residues 22–95) received a score of 2.610, which provides strong statistical support for the final model. Specifically, MetaMQAP predicted that the dimer core exhibits a root mean square deviation (RMSD) of ~4.2. That the dimerization refinement improved the model is indicated by the "monomer alone" receiving a PROQ score of 1.980 and a MetaMQAP-predicted RMSD of ~4.5. In the final step of model refinement, the intrinsically disordered termini were added with MODELLER (Sali and Blundell, 1993). Our final ELF4 model displayed in Fig. 2(D) and Supplementary Material, Figs. S2C-E represents a plausible structural fold for the ELF4/EFL family.

EFL functional conservation: to determine if the *EFL* genes are associated with circadian biology, suggestive of a broad conservation of function, we performed *EFL* expression analyses. In published microarray experiments, the *ELF4* expression profile differed from the *EFL*s (DIURNAL website; see Supplementary Material, Fig. S3A; note, *EFL1* was not included as a probe on the ATH1 array). Under continuous light (LL), *ELF4* cycled with high amplitude compared to *EFL2, EFL3*, and *EFL4*, which were expressed to lower maxima (see Supplementary Material, Fig. S3A). Additionally, the *EFL3* profile was neither diurnal nor circadian, and we found that *EFL1* expression was not circadian (see Supplementary Material, Fig. S3B). We characterized *efl* T-DNA insertion lines, but we were unable to detect significant circadian dysfunction in these lines (data not shown). Given the circadian expression of *EFL2* and *EFL4*, together with structural similarity to ELF4 (see Supplementary Material, Fig. S2A), we needed to test the circadian function among *EFL* sequences.

The *EFL* expression data indicated that *ELF4* circadian function could be conserved in a subset of *EFL* genes. To test this idea, we evaluated the ability of *EFL* to ectopically confer circadian rhythmicity in the absence of *ELF4*. This is plausible as *ELF4* is expressed to a higher level than the *EFL*s (see Supplementary Material, Fig. S3). *EFL* coding sequences from Arabidopsis and other species were fused to the *ELF4* promoter and transformed into *elf4-1* plants harboring the *CCA1:LUC* reporter. Three independent T2 complementation lines were screened under LL, and the results are summarized in Fig. 3 and Supplementary Material, Table S1. In these experiments, we found that *EFL1, InEU916969*, and *HvEU916968* could fully complement the *elf4* loss-of-function phenotype. *CCA1:LUC* was overtly rhythmic, with a similar rescue as that seen with the wild-type control (*ELF4p:ELF4*). In addition, the expression level of *CCA1:LUC* was similar to the control for these tested lines (Fig. 3). *EFL1, InEU916969*, and *HvEU916968* also restored *elf4* imprecision. The average ratio was about 93% rhythmic seedlings per line in contrast to the average of 70% of seedlings with rhythmicity for *elf4-1* (see Supplementary Material, Table S1).

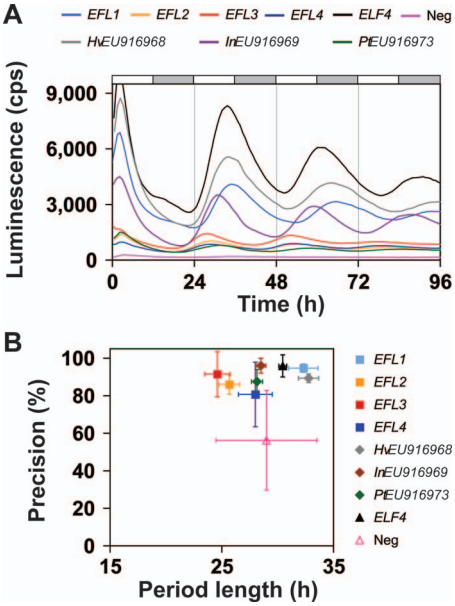

Figure 3. *EFL* complementation assay. *ELF4* circadian function is not conserved within all members of the DUF1313 fold. Seedlings harboring *CCA1:LUC* were entrained under 12L:12D cycles and released into LL. Lines are indicated by the identity of the insert. *ELF4*, *ELF4p:ELF4* positive control; Neg, negative control. See also Methods. Precision of rhythms were defined and measured as number of rhythmic seedlings with relative amplitude error (RAE)<0.5, as also listed in Table S1 (see Supplementary Material). Identity of sequences, as in Fig. 1. (A) *CCA1:LUC* profiles. Shaded boxes indicate subjective night. (B) Result of period analysis of profiles shown in (A); precision of population vs average period length. Error bars represent SD of two independent experiments.

EFL genes from outside the *ELF4* group were insufficient to restore *ELF4* function. *EFL2, EFL3, EFL4*, and *PtEU916973* all failed to complement *elf4-1* phenotypes (see Fig. 3 and Supplementary Material, Table S1). Although many T2 *CCA1:LUC* lines with ectopic *EFL2, EFL3, EFL4*, and *PtEU916973* expression displayed a modest statistical increase in the ratio of rhythmic seedlings under LL, no transgenic line restored *CCA1* amplitude. Taken together, the results from the *EFL* complementation assays are consistent with our phylogenetic results in that the more distantly related *EFL* genes (*EFL2* to *EFL4*) do not confer *ELF4* activity. In contrast, sequences within the *ELF4* group, such as

Table I. ELF4 TILLING. The ELF4 TILLING lines included in this study were named in the order the seeds arrived from the stock center. The site of the nucleotide change is listed according to the position in the genomic sequence.

Name	Stock No.	Mutation	Missense	Conservation	Residue change
elf4-215	N87889	G-153A	—	—	—
elf4-216	N86936	C-48T	—	—	—
elf4-201	N89610	G43A	E15K	Not conserved	Charge change (neg to pos)
elf4-202	N93293	G52A	E18K	Not conserved	Charge change (neg to pos)
elf4-207	N90524	C55T	Q19*	—	—
elf4-208	N87544	C68T	P23L	Not conserved	"Bending" to hydrophobic
—	N86474	G69A	—	—	—
—	N93422	G72A	—	—	—
elf4-209	N86619	G80A	W26*	—	—
elf4-203	N90093	C93T	R31W	Charge conserved globally (K, R)	Positive to hydrophobic
elf4-210	N88261	G94A	R31Q	Charge conserved globally (K, R)	Positive to hydrophilic
elf4-204	N91664	G103A	R34K	Charge conserved in subclade (K, R)	Both positive, size change
elf4-205	N91652	G133A	R44K	R conserved globally	Both positive, size change
elf4-211	N86681	C136T	S45L	Hydrophobicity conserved globally	Hydrophilic to hydrophobic
—	N91818	C164T	—	—	—
elf4-212	N86760	C178T	A59V	Not conserved, most are P	Both hydrophobic, size change
elf4-213	N90652	G222A	G74R	G conserved in subclade	Neutral to positive, size change
elf4-206	N89787	G340A	—	—	—
elf4-217	N88032	G370A	—	—	—
—	N89936	G434A	—	—	—
—	N87433	C553T	—	—	—

EFL1, InEU916969, and *HvEU916968*, are sufficient to provide ELF4 activity.

ELF4 TILLING

In order to functionally test our proposed ELF4 structure, we characterized a hypomorphic collection of *elf4* mutants that we obtained after a TILLING mutagenesis screen. Using this approach, we isolated ten alleles of *ELF4* with missense mutations in the coding region (see Supplementary Material, Fig. S4 and Table I).

Allelic strength: we generated transheterozygous *elf4* plants in order to test the allelic strength of the TILLING mutations. For this, we crossed *elf4-1* (Ws) harboring a *LUC* reporter (*CCA1:LUC* or *CCR2:LUC*, respectively) as a male to females of the *elf4* TILLING lines (Col-0). We note that for all F1 plants, one chromosome of a homologous pair will be entirely Ws and the other entirely Col-0. Thus, all lines were isogenic for genome-wide heterozygocity, except at the *ELF4* locus. We analyzed the F1 plants under LL, and the results are shown in Supplementary Material, Figs. S5A-C. When the two null alleles *elf4-207* and *elf4-1* were combined *in trans*, the *CCA1:LUC* expression dampened rapidly and the circadian rhythm was lost after ~48 h (see Supplementary Material, Fig. S5A). Furthermore, the mean level of *CCA1:LUC* was lower in *elf4-207*×*elf4-1* compared to Col-0×*elf4-1*, and this result confirmed the robustness of the assay. We generated F1 plants with the remaining *elf4* TILLING mutations. For plants containing *elf4-203* or *elf4-212* over the *elf4-1* null, we consistently found a short-period phenotype (see Supplementary Material, Fig. S5B and data not shown). Additionally, we found strong allelic strength (short period) for *elf4-203* and *elf4-212* in F1 plants harboring the *CCR2:LUC* reporter (see Supplementary Material, Figs. S5E and S5F). We note that loss of *LUC* rhythmicity (dampening) was only seen in *elf4-207*×*elf4-1* F1 plants (see Supplementary Material, Figs. S5A and S5D). All other allele crosses were similar to the Col-0×*elf4-1* results (data not shown), and this could reflect hybrid vigor. The F1 phenotypes thus indicated that *elf4-203* and *elf4-212* were relatively strong *elf4* alleles.

Diurnal gene expression in *elf4-207*: it was reported that *elf4-101* (null mutant in Col-0) had low *CCA1* and *LHY* expression, but unchanged *TOC1* levels during first subjective day under continuous red light (Kikis *et al.*, 2005). We sought to expand this clock-expression profiling in the equivalent Col-0 *elf4-207* allele, and first analyzed diurnal RNA samples taken from an LD regime. The diurnal phase of several clock genes (*CCA1, LHY, PRR9, PRR7*, and *GI*) was largely unchanged in *elf4-207*, indicating driven rhythms [Figs. 4(A)–4(E)]. This was in agreement with our previous characterization of *elf4* loss-of-function in Ws (McWatters *et al.*, 2007). However, we noted that *CCA1* and *LHY* displayed a low amplitude, *PRR7* was slightly derepressed in the night, and the *GI* expression, on average, was higher with lower kurtosis [Figs. 4(A), 4(B), 4(D), and 4(E)]. In addition, we found that the level of *TOC1* expression was elevated in

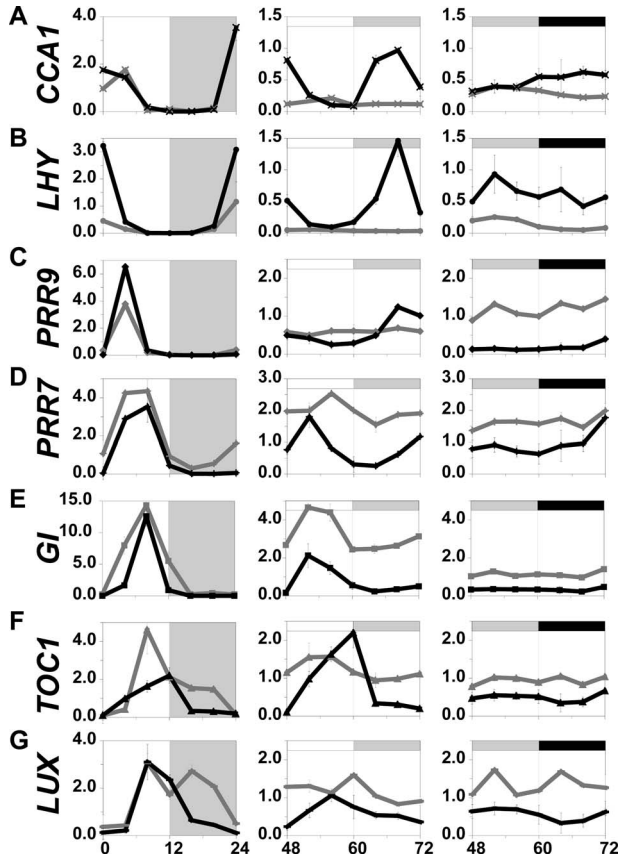

Figure 4. Clock gene expression in *elf4-207*. mRNA accumulation in *elf4-207* from (left panel) day seven under an LD cycle; (middle panel) the third day under LL; or (right panel) the third day in DD. The expression level is relative to *TUBULIN2 (TUB2)* and normalized to the average diurnal expression level in Col-0. Gray box indicates night-time for LD profiles, or subjective night or day in LL or DD assays. Gray and black curves represent *elf4-207* and Col-0, respectively. The *Y*-axes represent normalized gene expression and the *X*-axes are Time (hours). [(A) and (B)] Strongly attenuated expression of *CCA1* and *LHY* in *elf4-207*. (C) *PRR9*. LL average: *elf4-207*, 0.60; Col-0, 0.60; $P=0.98$. DD average: *elf4-207*, 1.18; Col-0, 0.18; $P=1.4\times10^{-6}$. (D) *PRR7*. LL average: *elf4-207*, 1.97; Col-0, 0.82; $P=0.0006$. DD average: *elf4-207*, 1.63; Col-0, 0.95; $P=0.002$. (E) *GI*. LL average: *elf4-207*, 3.21; Col-0, 0.77; $P=0.0002$. DD average: *elf4-207*, 1.14; Col-0, 0.34; $P=9.5\times10^{-7}$. (F) *TOC1*. LL average: *elf4-207*, 1.21; Col-0, 0.82; $P=0.27$. DD average: *elf4-207*, 0.94; Col-0, 0.50; $P=5.6\times10^{-6}$. (G) *LUX*. LL average: *elf4-207*, 1.16; Col-0, 0.60; $P=0.002$. DD average: *elf4-207*, 1.34; Col-0, 0.56; $P=8.2\times10^{-5}$.

the night phase [Fig. 4(F)]. We also analyzed the expression of the evening gene *LUX*, as this gene was hypothesized to have similar position in the circadian system as *TOC1* (Hazen *et al.*, 2005; Onai and Ishiura, 2005). Interestingly, the nighttime expression of *LUX* was also derepressed [Fig. 4(G)]. Thus, *ELF4* genetically functions as a repressor of *TOC1* and *LUX* transcript accumulation in the night phase of a diurnal cycle.

We next assayed the diurnal expression profiles of the ten *elf4* TILLING alleles. In these assays, we were unable to detect significant differences for the clock genes tested (data not shown). We conclude here that the LD cycle drives the circadian system in all our *elf4* hypomorphic alleles and likely masks any circadian dysfunction.

Circadian gene expression in *elf4-207*: to expand the expression profiling of *elf4-207*, we characterized circadian-gene expression under constant conditions, LL and darkness (DD). First, we confirmed that *elf4* loss-of-function in Col-0 led to arrhythmicity with low *CCA1* and *LHY* expressions, and a high level of *TOC1* under LL, similar to what we previously reported in Ws (Doyle *et al.*, 2002; McWatters *et al.*, 2007). Here, we found that the *TOC1* profile had reduced amplitude but was still moderately rhythmic [Fig. 4(F)]. This result differed slightly from the arrhythmicity and high transcript level of *TOC1:LUC* in *elf4-1*, and confirmed our earlier observation that an *ELF4* null allele in Ws is phenotypically more impaired than a null in Col-0 (McWatters *et al.*, 2007). We then assayed *LUX* expression in *elf4-207* and found a derepression and a shift in the peak phase of expression [Fig. 4(G)]. This result suggested that ELF4 also repressed *LUX* under LL. For *elf4-207*, the *TOC1* and *LUX* profiles in DD were similar to the result from the LL time course, except the mean level of both transcripts were relatively higher in DD than under LL [two- and threefold higher, respectively; Figs. 4(F) and 4(G)]. Thus, *ELF4* is required to sustain the circadian transcript accumulation of both *TOC1* and *LUX* without regard to the light environment, but *ELF4* has a significant role in the transcriptional repression of these two evening genes in DD.

We finished our *elf4-207* characterization by expression analysis of additional central-clock genes under constant conditions. Under LL, two genes in the "morning loop," *PRR9* and *PRR7*, dampened to peak and mean level of expression, respectively, in *elf4-207* [Figs. 4(C) and 4(D)]. Interestingly, this was not simply predictable from the mathematical model, as low *CCA1/LHY* expression should lead to the absence of *PRR9* and *PRR7* expressions (Locke *et al.*, 2006; Zeilinger *et al.*, 2006). In DD compared to under LL, the amplitude of the *PRR9* and *PRR7* expression was dramatically attenuated in Col-0 [most evident for *PRR9*; Figs. 4(C) and 4(D)]. Overall, we found that the derepression of *PRR9* and *PRR7* in *elf4-207* was more pronounced in DD than under LL. In *elf4-207*, *PRR9* was tenfold, and *PRR7* was twofold higher than in Col-0 in DD [Figs. 4(C) and 4(D)]. Regarding *GI*, under both LL and in DD, we found an increase in the *GI* expression level in *elf4-207* [Fig. 4(E)], although under LL, the peak time was unchanged. We conclude that *ELF4* controls the oscillations in the connected morning and evening loops (*CCA1/LHY-PRR9/PRR7* and *TOC1-Y/GI*, respectively) and, thus, *ELF4* has a double-entry point into the three-loop oscillator.

Modeling ELF4 as a dual factor in the clock: in order to relate the *elf4-207* expression phenotype to a current mathematical model of the circadian system, we modeled a pre-

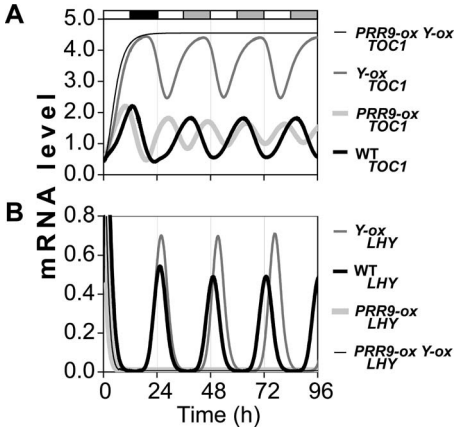

Figure 5. Mathematical modeling the circadian output of *PRR9-ox Y-ox*. Numerical simulation of the expression level of *TOC1* and *LHY* in *PRR9-ox*, *Y-ox*, and the double overexpression line, *PRR9-ox Y-ox*, compared to wild type, under one LD cycle and into LL. *PRR9* represents a compound activity of both *PRR9* and *PRR7*, and *Y* here was taken as *GI*, as in the three-loop model. The simulations were performed using the three-loop model. (A) The expression level of *TOC1* dampens high and becomes arrhythmic only following the overexpression of both *PRR9* and *Y*. (B) Overexpression of both *PRR9* and *Y* results in elimination of the *LHY* expression. Note that *LHY* was found at the base-line in the *PRR90x* and *PRR9-ox Y-ox* models.

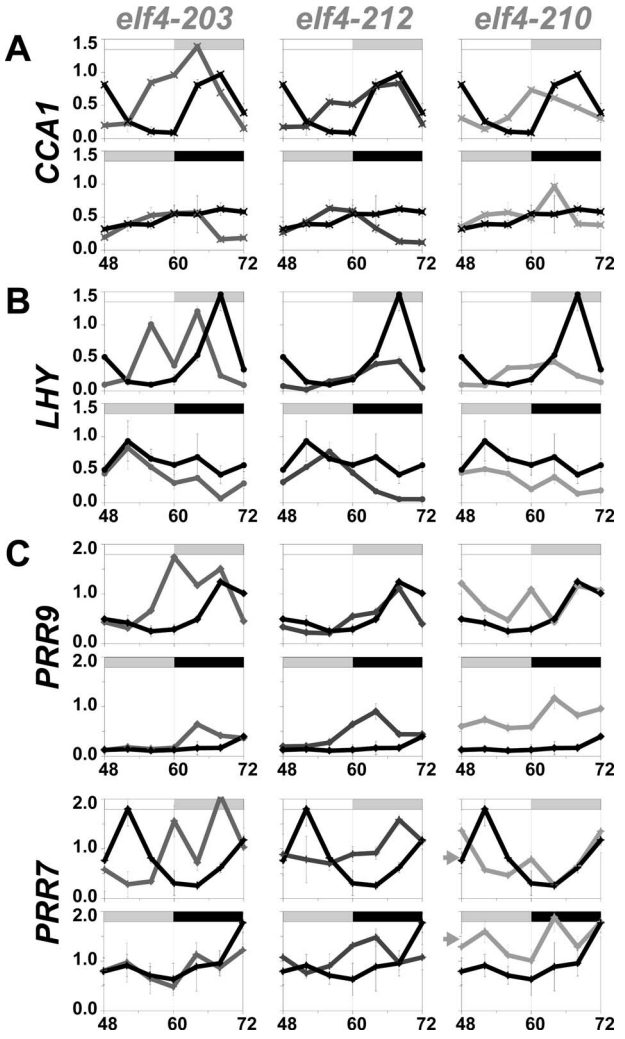

Figure 6. Morning-clock gene profiles in some *elf4* alleles. Clock transcript accumulation in *elf4-203*, *elf4-212*, *elf4-210* (all in respective gray), and Col-0 (black) under the third day under LL (lower panel), or the third day in DD (upper panel). Normalized data and Col-0 data as in Fig. 4. The *Y*-axes represent normalized gene expression and the *X*-axes are time (hours). (A) *CCA1*; (B) *LHY*; (C) *PRR9*. Mean expression values as follows. *elf4-203*: LL, 0.89; DD, 0.29; $P=0.03$. *elf4-212*: LL, 0.49; DD, 0.44; $P=0.75$. *elf4-210*: LL, 0.88; DD, 0.78; $P=0.52$. Col-0: LL, 0.60; DD, 0.18; $P=0.03$. (D) *PRR7*. Arrowheads indicate the derepression phenotype in *elf4-210* (LL, 0.78; DD, 1.43; $P=0.008$).

dicted connection point of *ELF4* in the clock. As arrhythmic *elf4-207* had an unpredictably high *PRR9* and *PRR7*, as well as an unpredictably high *GI* expression, we modeled the effect of the elevated *PRR9/7* expression (here simplified as one gene, *PRR9*), in addition to the increased expression of the hypothetical gene *Y*, which is partially explained by *GI* (Locke *et al.*, 2006). We elevated an arrhythmic *PRR9/7* at two-fold compared to the wild type, and also *Y* at three-fold the wild type, as the experimentally determined transcript levels of *PRR9/7* and *GI* in *elf4-207* under LL [Figs. 4(C)–4(E)]. Interestingly, we found that the *elf4* null phenotype was mimicked by this simultaneous misexpression of *PRR9* and *Y* using the three-loop model (Locke *et al.*, 2006) (Fig. 5). We note that in this test of the model, the expression of *TOC1* was only found to dampen when both *PRR9* and *Y* had an elevated expression and not in either single misexpression model [Fig. 5(A)]. The expression of *LHY* (that represents both *CCA1* and *LHY*) was rhythmic in the single *Y-ox* model, whereas it was at background levels in the *PRR9-ox* single and the double misexpression model [Fig. 5(B)]. Thus, when the three-loop model has simultaneously high *PRR9/7* and high *Y*, it results in oscillator arrest with an attenuation of *LHY* and an elevation of *TOC1*. This is qualitatively similar to what we observed experimentally in *elf4-207* [Figs. 4(A), 4(B), 4(F), and 4(G)], and collectively, these modeling results support our idea that *ELF4* has dual input pathways to the loop network; *ELF4* functions within the clock to repress *PRR9* and *PRR7*, as well as to repress *GI*.

Circadian gene expression in hypomorphic *elf4* lines: we predicted *elf4-203* and *elf4-212* to have the largest phenotypic changes in the clock-gene expression, based on the transheterozygous test of allelic strength (see Supplementary Material, Fig. S5). Here, we could define *elf4-203* to be the strongest hypomorphic *elf4* allele in the TILLING panel (Figs. 6 and 7). This allele had a 4–8 h early phase of all genes under LL and most genes in DD, indicative of a short circadian period (Figs. 6 and 7). We note that compared to *elf4-207, elf4-203* was found to display wild-type mean expression levels of the morning genes, but an increase in *LUX* transcript [*LUX* normalized levels at 2.26 in *elf4-203* and

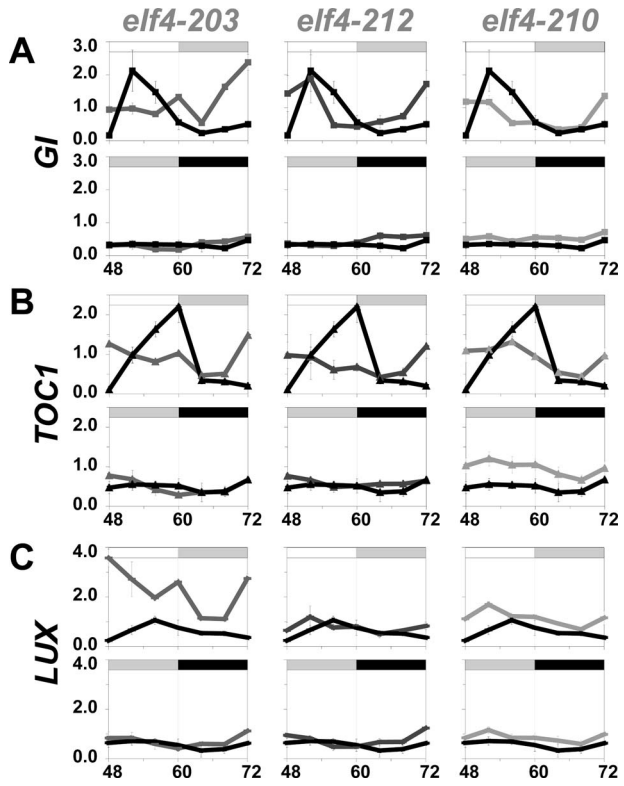

Figure 7. Evening-clock gene profiles in some *elf4* alleles. Clock transcript accumulation in *elf4-203*, *elf4-212*, *elf4-210* (all in respective gray), and Col-0 (black) under the third day under LL (lower panel), or the third day in DD (upper panel). Normalized data and Col-0 data, as in Fig. 4. (A) *GI*; (B) *TOC1*; and (C) *LUX*. The *Y*-axes represent normalized gene expression and the *X*-axes are time (hours).

0.59 in Col-0; $P=0.002$; Fig. 7(C)]. Under both LL and DD, the phase of the evening-gene expression (*GI, TOC1*, and *LUX*) was similar in *elf4-203*, and this was not true in the wild type, suggesting an uncoupling of the evening loops in *elf4-203* (Fig. 7).

The second strong *elf4* allele, *elf4-212*, had rhythms that were less phase-shifted than *elf4-203*. The ~4 h phase shifts were pronounced for *CCA1, PRR7*, and *GI* under LL (Figs. 6 and 7). In addition, *elf4-212* had low amplitude of *LHY* [Fig. 6(B)]. For *PRR9*, the expression level was not low in DD in *elf4-212* compared to in *elf4-212* under LL [*PRR9* normalized levels in *elf4-212* under LL at 0.49, and in DD, at 0.44, $P=0.75$; Col-0 under LL at 0.60, and in DD at 0.18; $P=0.03$; arrows in Fig. 6(C)], suggesting failure of *elf4-212* to repress *PRR9* levels in DD.

We analyzed *elf4-210*, which is mutationally related to *elf4-203*. The mutation in *elf4-210* (R31Q) affected the same site, as in *elf4-203* (R31W; Table I), but with a more conservative change in the substituted encoded amino-acid. Similar to *elf4-203*, all clock-gene expressions in *elf4-210* displayed a phase change under LL (Figs. 6 and 7). Similar to *elf4-212, elf4-210* had a derepression phenotype of *PRR9* in DD, where the level was the same in DD as under LL [*PRR9* normalized levels in *elf4-210* under LL at 0.88, and in DD at 0.78, $P=0.52$; Fig. 6(C)]. Unlike *elf4-203* and *elf4-212, elf4-210* had a dark-phenotype with respect to the mean level of *PRR7*. In *elf4-210*, the average level of *PRR7* was elevated in DD compared to *elf4-210* under LL [*PRR7* normalized levels for *elf4-210* under LL at 0.78, and in DD at 1.43; $P=0.008$; arrowheads in Fig. 6(D)]. Regarding evening-gene expression in *elf4-210*, the level of *TOC1* was elevated in DD (*TOC1* normalized levels in *elf4-210* at 0.96, and Col-0 at 0.50, $P=0.0002$), and the *LUX* expression displayed a phase shift [Figs. 7(B) and 7(C)]. Collectively, from the analysis of the *elf4-203*, *elf4-212*, and *elf4-210* expression profiles, it appeared that *ELF4*-mediated control of clock phase and amplitude involves the regulation of *LUX* levels under LL, and separately the regulation of *TOC1*, and *PRR9* and *PRR7* expression levels in DD. This observation is consistent with our hypothesis that *ELF4* has a dual entry point to the clock (Fig. 5).

We define *elf4-204, elf4-202*, and *elf4-201* as medium hypomorphic alleles. All three alleles had derepression phenotypes regarding *PRR9* and *PRR7* expression, where the level was, respectively, the same or higher in DD than under LL (e.g., *PRR7* in *elf4-204* under LL at 0.40, and in DD at 1.48; $P=0.0002$; see Supplementary Material, Fig. S6). In addition, *elf4-204* had the same expression level of *GI, TOC1*, and *LUX*, respectively, under LL, as in DD (see Supplementary Material, Fig. S6). Both for *elf4-201* and for *elf4-202*, *TOC1* levels were high in DD, and *LUX* was high under LL (see Supplementary Material, Figs. S6B and S6C). Regarding *GI*, *elf4-201* had a high expression and *elf4-202* had a low expression under LL (see Supplementary Material, Fig. S6A). We note here that the mean expression levels *CCA1* and *LHY* were normal in *elf4-204*, *elf4-201*, and *elf4-202* (data not shown). Thus, the phenotypes of three medium hypomorphic alleles indicate that *ELF4* can affect the morning and evening loops without a change in the expression levels of *CCA1* and *LHY*.

Our analysis of the four remaining weak alleles, *elf4-208*, *elf4-205*, *elf4-211*, and *elf4-213* confirmed a trend of incomplete *PRR9* and *PRR7* repressions in DD (see Supplementary Material, Fig. S7). In addition, *elf4-205*, *elf4-211*, and *elf4-213* had similar low expression of *GI* under LL, as in DD (Fig. 8(E) and data not shown), and similar constant levels of *TOC1* were present in *elf4-208*, *elf4-205*, and *elf4-211* (Fig. 8(E) and data not shown).

Surface model of the ELF4 dimer: we integrated our data from the above experiments into one rational explanation of the structure-function relation of ELF4. This served to connect its phylogeny and positions of sequence conservation with both our *elf4* expression phenotypes and the positioning of the amino acids, whose missense replacement result in a phenotype, to the structural model of ELF4. Furthermore, this allowed us to assess the relative orientation of amino ac-

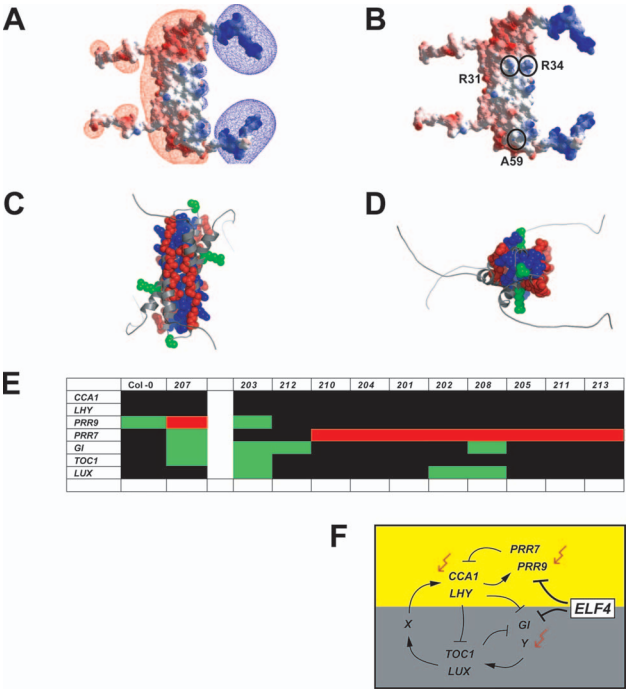

Figure 8. Integrated models of ELF4 at a structural and functional level. (A) The strong missense alleles of ELF4 are predicted to affect charge distribution on the ELF4 structure. The distribution of charged residues surrounding the proposed ELF4 dimer are "opposite" in such a way that only one side of the dimer is positively charged and vice versa. Blue color depicts positive charged surface and red depicts negatively charge surface. Note that the ELF4 dimer has a polar ionic distribution. (B) Position of the residues affected in *elf4-203* (R31), *elf4-210* (R31), *elf4-204* (R34), and *elf4-212* (A59) on the surface model. (C) ELF4 dimer, side view. The globally conserved residues are colored in blue. Red residues are the amino acids that are conserved in the ELF4 subgroup (see also Supplementary Material, Fig. S1). R31, R34, and A59 described in (B) are highlighted in green. (D) ELF4 dimer, end view. Colors as in (C). (E) Summary of all expression data from Figs. 4, 6, and 7 and Supplementary Material, Figs. S6 and S7. The coloring represents gene expression (mean level) that is not changed (black), increased (red), or decreased (green) as compared between conditions (LL and DD) within a given genotype. (F) A genetic model of *ELF4* in the three-loop model of the circadian system. *ELF4* repress both *PRR9*/*PRR7* (the morning loop) and *GI* in the evening loop. Red arrows indicate points of light input.

ids for those sequences conserved within the DUF1313 domain, in general, and the ELF4/EFL1 clade, in particular, and to comprehend the surface residues that were mutagenic targets. First, we assessed the surface potential of the ELF4 dimer shown in Fig. 2. Strikingly, the dimer interface resolves into a nonsymmetrical electrostatic distribution. Shown in Fig. 8(A), the dimer is predicted to have a positively charged surface across one-half of the dimer and a negatively charged surface over the other. This remarkable electrostatic asymmetry overlaps with the strong missense alleles, *elf4-203* and *elf4-210*, and the medium allele *elf4-204* [R31 and R34, respectively; Fig. 8(B)]. Note that none is positioned within the negative field. Furthermore, the mutations at R31 (R31W for *elf4-203* and R31Q for *elf4-210*) led to loss of charge, whereas the mutation at R34 (R34K for *elf4-204*) resulted in a conservative change that continues to have a positive charge. The second strong allele *elf4-212* encodes the A59V mutation, which is at the hinge and just outside the charged fields [Fig. 8(B)].

We sought to characterize the phylogenetic conservation of residues over the ELF4 dimer structure. Highlighting the 15 residues that are conserved across all ELF4 and EFL sequences, relative to ELF4 from Arabidopsis; these residues are Q37, L40, N43, R44, L46, I47, N53, N65, V66, I69, E71, N73, N75, V79, and Y83 (see Supplementary Material, Fig. S1); these are colored in blue in Figs. 8(C) and 8(D); and Supplementary Material, Fig. S2D. Similarly, highlighting the 12 residues that are conserved across ELF4/EFL1, but distinct from the other EFLs, these residues are R34, R42, Q48, Q49, D52, Q55, Q70, G74, S77, N81, S84, and S87 (see Supplementary Material, Fig. S1); these are colored in red in Figs. 8(C) and 8(D), and Supplementary Material, Fig. S2D. Finally, highlighting the strongest TILLING alleles, these residues are R31 and A59, and are colored in green in Figs. 8(C) and 8(D), and Supplementary Material, Fig. S2D. What can be seen in Figs. 8(C) and 8(D), and Supplementary Material, Fig. S2D is a distinct and nonrandom sorting of features. The residues conserved throughout the DUF1313 family position within the interior of the dimer. The residues conserved only in the ELF4/EFL1 clade are positioned away from the dimer interface and, furthermore, these side chains have an outward directionally toward the solvent. The amino acids that were encoded deviant in *elf4-203*/*elf4-210* (R31) and *elf4-212* (A59) also face away from the interior of the ELF4 structure [see Figs. 8(C) and 8(D) and Supplementary Material, Fig. S2D]. Collectively, these insights lend strong support for our proposed ELF4 structure to the extent that the model likely has correctly positioned the directionality of the side chains. Our results also implicate that the evolving face of ELF4 is within an asymmetric electrostatic exterior of the surface.

DISCUSSION

Using loss-of-function alleles of *elf4*, it was previously found that *ELF4* is important for clock precision and, in addition, gain-of-function studies supported *ELF4* as a repressor of clock periodicity (Doyle *et al.*, 2002; Kikis *et al.*, 2005; McWatters *et al.*, 2007). Here, we analyzed the *ELF4* structure-function relationship using a fully integrated and interdisciplinary approach combining phylogenetics, structural modeling, biochemistry, systems modeling, and molecular genetics. These integrated studies first led us to define the ELF4 functional domain and then to predict and biochemically confirm that ELF4 forms a homodimer with α-helical composition. We then extended these analyses by applying a TILLING mutagenesis screen over the *ELF4* lo-

cus, where we isolated a new *elf4* mutant collection consisting of missense alleles. The mutant phenotypes of the *elf4* hypomorphic alleles revealed that *ELF4* couples the three loops of the circadian system and, furthermore, that *ELF4* connects to the repression of *GI*, *TOC1*, and *LUX* and specifically to dark-repression of *PRR9* and *PRR7*. This supports the placement of *ELF4* within two nodes of the oscillating mechanism. Additionally, we note that the hypomorphic alleles confirmed findings that *ELF4* is a repressor of the speed of the clock. The structural map of the hypomorphic series allowed us to define the exterior area of the ELF4 dimer critical for function. Interesting, this interface is in a distinct electrostatic surface. Collectively, we conclude from these studies that *ELF4* represses morning, as well as evening loops of the circadian system and, in this way, it entrains the clock.

ELF4 phylogeny

Our multiple alignment of ELF4 and EFL sequences identified the central/core ELF4 (DUF1313) domain, which was highly conserved. Furthermore, two "tails" could be determined and these, based on the degree of conservation, were predicted to be less important for the ELF4 structure (see Supplementary Material, Fig. S1). Our phylogenetic results identified *ELF4* and *EFL1* as members of sister gene lineages in eudicots and that the two sequences from *Populus trichocarpa* (VTK45573 and YFS280019) were members of the *ELF4* and *EFL1* lineages, respectively (Fig. 1). The available sequences from monocots (all from grasses) were more closely related to each other than to either of these genes. Arabidopsis *EFL2* to *EFL4* formed a clade by themselves and their relationships with *ELF4/EFL* sequences in other species remains ambiguous. Several species have *ELF4/EFL* sequences that appear to be distantly related to each other and that also are not closely related to sequences from other species. Together, the evolution of *ELF4/EFL* genes is characterized by a dynamic history of gene duplication and loss. Further sampling will provide a more comprehensive description of *ELF4* versus *ELF* evolution.

In accordance with our phylogenetic analyses, modeling of ELF4 and many EFL sequences revealed similar predicted structures for these proteins (see Fig. 2 and Supplementary Material, Fig. S2A). We note that the predictions were computationally independent and, thus, the consistency of the data led us to conclude that the general fold of ELF4 had been discovered. All ELF4/EFL sequences were predicted to consist of two α-helices and we therefore tested whether this fold contained *ELF4/EFL* functional activity. However, by ectopic complementation assays, of expression of *ELF* genes in the *elf4* minus background, we did not find support that the DUF1313 fold was globally conserved for ELF4 function (Fig. 3). We conclude that the difference in *ELF4p:EFL* complementation ability was consistent with the groupings in the *ELF4* phylogenetic tree.

Missense mutations in *elf4*

The sequence conservation in the ELF4 family (see Supplementary Material, Fig. S1) provided a basis for the prediction of the mutant phenotypes of the *elf4* TILLING alleles [Table I and Figs. 8(A) and 8(B)]. Specifically, *elf4* mutations could be divided into two classes based on the encoded difference in charged residues. Both *elf4-201* (E15K) and *elf4-202* (E18K) (see Supplementary Material, Fig. S6) encoded a charge change, but as both E15 and E18 were located in the amino-terminal tail (little sequence conservation), *elf4-201* and *elf4-202* were not predicted to have strong phenotypes. The P23L substitution in *elf4-208* (see Supplementary Material, Fig. S7) changed a proline residue that was not conserved, and because this location was at the end of the tail, no major mutation effect was expected. The more centrally positioned mutations were interesting because the central region had the highest number of conserved ELF4-residues and encoded the two α-helical folds (Fig. 2 and Supplementary Material, Fig. S1). In two mutants, *elf4-203* (R31W) and *elf4-210* (R31Q) (Figs. 6 and 7), the same encoded arginine residue was changed; more importantly the positive charge was globally conserved at this position, hence, the replacement of arginine with an uncharged amino-acid suggested detectable phenotypes in *elf4-203* and *elf4-210*. Both *elf4-204* (R34K) and *elf4-205* (R44K) (see Supplementary Material, Figs. S6 and S7) encoded no change in charge and affected only the size of the residue. The location of R34 and R44 in the α-helical fold indicated that a size change would confer a phenotype. Additionally in *elf4-205*, a conserved arginine was replaced with lysine and this coincided with a phenotype. Mainly hydrophobic residues are found at residue 45 in the ELF4 phylogeny, therefore, a phenotypic effect from the replacement of serine with lysine in *elf4-211* (see Supplementary Material, Fig. S7) was unexpected. The A59V change in the *elf4-212* mutant (Figs. 6 and 7) affected an alanine residue that was not conserved; however, because this alanine was located in the hinge region between the two α-helical folds, an amino-acid size change at this position implied a mutant phenotype. The most C-terminal residue affected in the *elf4* collection was G74 in *elf4-213* (see Supplementary Material, Fig. S7). This glycine was conserved in the ELF4 subgroup, suggesting a structural role. The replacement of this glycine with arginine introduced a positive charge at this position, in addition to a size change, which was predicted to disorder the ELF4 surface and the α-helical fold. Collectively, the mutations correlated with the expected phenotypic analysis of the mutants when considered in a structural and phylogenetic context.

To confirm the *ELF4*-specificity of the *elf4* TILLING mutations, we analyzed transheterozygous plants (see Supplementary Material, Fig. S5)—a classical genetic tool developed in 1941 (Lewis, 1941)—and an approach used previously in the analysis of TILLING mutants (e.g., Bao *et al.*, 2004; Enns *et al.*, 2005; Mizoi *et al.*, 2006). We used

elf4-1 (Ws) as the null allele because this is a true null (the *ELF4* locus is deleted; Doyle *et al.*, 2002). In this way, using the two reporters, *CCA1*:*LUC* and *CCR2*:*LUC*, we proved that the short-period phenotypes of the strong *elf4-203* and *elf4-212* alleles were caused by reduced *ELF4* function (see Supplementary Material, Fig. S5). The transheterozygous plants with the other *elf4* TILLING alleles did not display phenotypes for *LUC*-derived bioluminescence rhythms (data not shown), which perhaps indicate heterosis in the hybrid Col-0 × Ws genome.

ELF4 dimer

The coiled-coil-like dimer structure of ELF4 suggested that this fold could serve as an interaction platform. Noting that the ELF4 dimer model has a predicted asymmetric electrostatic surface, this implicated a molecular activity for this protein. The 12 residues conserved in ELF4/EFL1 subgroup were found to be surface-facing and, thus, would be solvent-exposed [Figs. 8(C) and 8(D)]. As these surface residues were conserved, we suggest that they could serve to direct salt-bridge interactions to a protein-interaction target. That residue 31 is positioned within the positive electrostatic cluster could implicate this particular half of the dimer as the surface of such an activity. We found 15 fully conserved residues in the DUF1313 family that appeared within the interior of the dimer structure. This could implicate that these residues are structural for the α-helical fold and for the strong homodimer interaction. Collectively this leads us to hypothesize that ELF4 functions as a molecular "key." Here, the activity of ELF4 would be to bind a target protein and such binding would activate that receptor molecule whilst assembled to the ELF4 effector ligand. Future efforts will explore if such a molecular activity exists for ELF4. If so, understanding this activity would move towards a biochemical definition of how ELF4 directs the different arms of the circadian clock.

Clock gene expression in *elf4*

The *elf4* null mutant has dramatically attenuated *CCA1* and *LHY* expression (Doyle *et al.*, 2002; Kikis *et al.*, 2005). More recently, we have shown that *TOC1* expression is elevated in the *elf4* null (McWatters *et al.*, 2007). In our attempt to extend the network of *ELF4* action, we analyzed the expression of clock genes in *elf4-207*, an *elf4* null allele with an apparently identical phenotype to *elf4-101* (Khanna *et al.*, 2003). New to this network placement of *ELF4*, we found that in *elf4-207*, *PRR9* and *PRR7* were expressed at a high level relative to the wild type, and this was most obvious in DD [Figs. 4(C) and 4(D)]. Recalling that low *CCA1*/*LHY* levels are seen in *elf4* (Kolmos and Davis, 2007), we interpret the elevated *PRR9* and *PRR7* expression as a direct effect of ELF4 because according to the three-loop model the *PRR9* and *PRR7* genes would be inactivated in the absence of the *CCA1* and *LHY* expression (Locke *et al.*, 2006; Zeilinger *et al.*, 2006). How *TOC1* feeds back on *CCA1*/*LHY* expression has long been questioned, and this was one reason why a hypothetical gene *X* was incorporated into mathematical models (Locke *et al.*, 2006; Zeilinger *et al.*, 2006). Until *X* is identified, its relation to the *ELF4* function is an open question. Furthermore, *TOC1* has also been placed in a negative feedback with *Y*/*GI*. The high levels of both *GI* and *TOC1* expression in *elf4-207* [Figs. 4(E) and 4(F)] conform to the notion that *GI* can only be a partial constituent of *Y* (e.g., Ito *et al.*, 2008; Locke *et al.*, 2006; Martin-Tryon *et al.*, 2007).

Our comparative analysis of the *PRR9* and *PRR7* expression in the *elf4* collection led us to the conclusion that *PRR7* expression in DD is more sensitive to changes in *ELF4* activity. Interestingly, this derepression phenotype was also detectable under an LD cycle, where *PRR7* was elevated in the night in *elf4-207* compared to Col-0 [Fig. 4(D)]. Through conserved *cis*-regulatory elements *CCA1* and *LHY* are activators of *PRR7* transcription (Farre *et al.*, 2005); however, in DD we did not find concomitant derepression phenotypes of *CCA1* and *LHY* in the *elf4* alleles. The *PRR9* derepression in the strong *elf4* alleles was in agreement with earlier findings that *PRR9* expression is light-induced and that *elf4* gates light input (Ito *et al.*, 2007; McWatters *et al.*, 2007). Collectively, the gated control of *PRR9* and *PRR7* in DD seemed to have been uncoupled in our *elf4* hypomorphic alleles.

LUX is an evening-expressed transcription factor believed to function at the same phase as *TOC1* and *ELF4* (Hazen *et al.*, 2005; Onai and Ishiura, 2005). We found that the null *elf4-207* had elevated *LUX* expression [Fig. 4(G)]. Under LD cycles, *elf4-207* displayed a derepression phenotype for *TOC1* and *LUX* expressions. This difference was also detected in the expression of *GI* in *elf4-207*, which could indicate indirect action of *elf4* on *TOC1 via GI/Y*. Furthermore, *ELF4* likely has a similar position in a (separate) negative feedback loop with *LUX* because the *lux* mutant has increased *ELF4* transcript accumulation (Onai and Ishiura, 2005). Therefore, we predict *LUX* as a candidate for a component of *Y*, a hypothesis that remains to be tested.

Our analysis of *elf4-207* suggests that *ELF4* has two inputs, at both *PRR9*/*PRR7* and *GI*/*LUX* expression, respectively, and that it functions to repress the (light-induced) expression of these "entry" points [Fig. 8(F)]. This is in agreement with the three-loop model in that a high rate of *GI* expression would lead to a high *TOC1* transcript level. This activity would be in addition to our earlier findings that ELF4 is an entrainment factor (Kolmos and Davis, 2007; McWatters *et al.*, 2007). Our use of the mathematical three-loop model confirms the plausibility of this hypothesis because the simultaneous overexpression of *PRR9* and *GI* resulted in an arrhythmic clock with peak transcript levels of *TOC1* and the absence of *LHY* expression (Figs. 4 and 5).

We appreciate here previous data in support of the above

conclusion: both in the *cca1 lhy toc1* triple mutant and in *elf4*, the clock arrests ~24 h after release from entrainment (Ding *et al.*, 2007; McWatters *et al.*, 2007), indicating a similar disruption of the loops in *cca1 lhy toc1*, and *elf4*. Hence, it is rational that *ELF4* targets more than one loop of the network. Furthermore, *elf4* is one of the few clock single mutants, apart from *lux* and *early flowering 3* (*elf3*) (Hazen *et al.*, 2005; Hicks *et al.*, 1996), that is arrhythmic (neither *LUX* nor *ELF3* has yet been placed into the three-loop model), and this could further support our idea that *ELF4* has more than one entry point to the clock [Fig. 8(F)]. This dual role could explain *ELF4* function in entrainment in that the clock has a different resetting response to dawn and dusk cues (Millar, 2003).

As light is the major entrainment signal (*Zeitgeber*), the circadian network likely runs differently under LL than in DD. The results of our expression analyses of the *elf4* hypomorphic alleles indicate that the circadian system is more compromised in DD than under LL, as phenotypes were more readily detected in DD in the weakest of the *elf4* missense mutants. In DD, many core-clock gene transcripts in *elf4* are not repressed as they are in the wild type. It is under such conditions that the *ELF4* function appears most critical. The mathematical modeling of the circadian system was primarily based on the interconnected-loop model under LD entrainment and with continuous light input (Locke *et al.*, 2006; Zeilinger *et al.*, 2006). Future expansions of the model with *ELF4* will help to answer the question about difference in circadian-loop performance under LL vs in DD.

MATERIALS AND METHODS

Nomenclature

In this study, *ELF4-LIKE* (*ELF4-L*, for *ELF4* homolog, (Khanna *et al.*, 2003) was renamed *EFL* (for ELF4-LIKE) with the numbering as in Khanna *et al.* (2003); *EFL1* is At2g29950; *EFL2* is At1g72630; *EFL3* is At2g06255; and *EFL4* is At1g17455. The *EFL* gene name is in accordance with the guidelines for Arabidopsis gene nomenclature (Meinke and Koornneef, 1997).

EST clones

Candidate orthologous sequences were identified using BLAST (Altschul *et al.*, 1990) in GenBank and genome databases (DOE Joint Genome Institute, Physcobase, and TIGR). EST clones were obtained from the respective investigators, and these were fully sequenced. Contigs were assembled using the GCG package (Wisconsin Package, Genetics Computer Group, Madison, WI), and the consensus sequence was determined.

The sequenced EST clones in this study have been deposited in GenBank (http://www.ncbi.nlm.nih.gov): EU919439 (*Am*), EU916962 (*Bv*), EU916963 (*Cs*), EU916964 (*Ga*), EU916965 (*Ga*), EU916966 (*Ha*), EU916967 (*Ht*), EU916968 (*Hv*), EU916969 (*In*), EU916970 (*Sl*), EU916971 (*Ls*), EU916972 (*Mt*), EU916973 (*Pt*), EU916974 (*Sb*), EU916975 (*So*), EU916976 (*So*), EU916977 (*St*), EU916978 (*Zm*), EU916979 (*Mp*), EU916980 (*Gm*), and EU916981 (*Lc*).

The Arabidopsis Genome Initiative locus codes for the genes discussed in this paper are as follows: *CCA1*, At2g46830; *EFL1*, At2g29950; *EFL2*, At1g72630; *EFL3*, At2g06255; *EFL4*, At1g17455; *ELF4*, At2g40080; *GI*, At1g20620; *LHY*, At1g01060; *LUX*, At3g46640; *PRR7*, At5g02810; *PRR9*, At2g46790; and *TOC1*, At5g61380.

Phylogeny

The derived protein sequences of the sequenced ESTs, and additional partial EST sequences from the public databases, were compared using the multiple alignment tool CLUSTALX (Thompson *et al.*, 1994) and processed with BOXSHADE 3.21 (http://www.ch.embnet.org/software/BOX_form.html). The derived sequences for the predicted ELF4/EFL proteins are intentionally left un-named so that nomenclature projects can define the name relative to the respective plant community.

Phylogenetic analyses of a 204-nucleotide alignment were performed to assess homology of ELF4 and related sequences from angiosperms. Bayesian phylogenetic analyses were performed using MRBAYES 3.1.2 (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003). Two independent analyses were run, where each consisted of 2,000,000 generations with sampling every 1,000 generations. Maximum likelihood bootstrapping was performed using GARLI 0.951 (Zwickl, 2006). Likelihood parameters were estimated on a parsimony tree, using Paup* (Swofford, 2002). Maximum parsimony bootstrap analyses were performed using Paup* (Swofford, 2002) with 1,000 replicates, a heuristic search strategy of ten random sequences addition, tree-bisection-reconnection swapping, and saving multiple most parsimonious trees.

Plant materials

The *elf4-1* mutant (Ws) has been described (Doyle *et al.*, 2002). The luciferase lines used were Ws *CCA1*:*LUC* and Ws *CCR2*:*LUC* (Doyle *et al.*, 2002). T-DNA insertion lines were *efl1-2* (Col-0, SALK_135613), *efl2-1* (Ws, FLAG_198A02), *elf3-1* (Ws, FLAG_140E10), *efl3-2* (Col-0, SALK_009170), *efl3-3* (Col-0, SALK_092662), *efl3-4* (Col-0, SALK_078416), *efl4-1* (Col-0, SALK_084137), *efl4-2* (Col-0, SALK_058067), and *efl4-3* (Col-0, SAIL_837_C07), which were all obtained from NASC. *efl1-1* (Ws) and *efl4-3* (Ws) were provided from an initial study on *EFL* genes (Doyle, 2003; Krysan *et al.*, 1999).

TILLING was performed on the entire coding sequence of *ELF4*, and this region extended 880 bp. The primers used in the TILLING screen were ELF4-fwd: 5′-CCA ATC ACT TCA CAG CTT CAC TCA CG-3′, and ELF4-rev: 5′-TGC AAC AAT CTA ACC ACA AGC CTT CA-3′. In total, 21

new DNA lesions in the *ELF4* locus (Table I) were identified. The M3 TILLING plants were backcrossed three times to Col-0 wild type. Cleaved amplified polymorphims (CAPS)/derived cleaved amplified polymorphic sequences (dCAPS) markers for the TILLING point mutations are listed in the Supplementary Material, Table S2. The homozygous mutant BC3-F2 plants were confirmed by sequencing the affected gene region.

TILLING lines included in this study, with stock-center codes in parentheses: *elf4-201* (N89610), *elf4-202* (N93293), *elf4-203* (N90093), *elf4-204* (N91664), *elf4-205* (N91652), *elf4-206* (N89787), *elf4-207* (N90524), *elf4-208* (N87544), *elf4-209* (N86619), *elf4-210* (N88261), *elf4-211* (N86681), *elf4-212* (N86760), *elf4-213* (N90652), *elf4-215* (N87889), *elf4-216* (N86936), and *elf4-217* (N88032). No homozygous F2 plants were found for lines *elf4-206* and *elf4-217*, and for *elf4-211* and *elf4-212* the segregation pattern differed between seed batches from the same generation suggesting presence of linked lethal mutations in the backcrossed genomes. Transheterozygous (F1) plants were obtained by crossing *elf4-1*, using pollen from *elf4-1* homozygous for the *CCA1*:*LUC* or *CCR2*:*LUC* reporter genes to the *elf4* TILLING lines and Col-0. Multiple F1 seed were used, and F1 data presented represent the analysis of two biological replicates.

Complementation of *elf4-1*

The 35S promoter fragment of the binary vector pJawohl (Kan^R, gift from Bekir Ulker, MPIZ) was replaced with the *ELF4* promoter (amplified with primers 4pf-AscI, 5′-TAG TAA GGC GCG CCC TCA TGA TTT CCT GCG GTA ATT ATC T-3′; and 4pr-ClaI, 5′-TAC CGG ATC GAT AAT AAT TTT TAA TTG TGT TTT TCT CTC T-3′) using *Cla*I and *Asc*I sites to create pJawohl/ELF4p. Subsequent restriction with *Asc*I and *Spe*I enabled exchange of the promoter-Gateway cassettes of pJawohl/ELF4p and pLeela (Basta^R, gift from Marc Jakoby, MPIZ), to give pJalee4. Arabidopsis *ELF4* and *EFL* genes were amplified from genomic DNA (Ws ecotype). *Ipomoea*, *Hordeum*, and *Pinus* genes were amplified from EST plasmids.

The following primers (with appropriate Gateway overhangs, GW) were used for amplification: attB1-ELF4, 5′-GWF-ATG AAG AGG AAC GGC GAG ACG AAA-3′; attB1-EFL1, 5′-GWF-ATG GAA GCA TCG AGA AAT CGA TCG-3′; attB1-EFL2, 5′-GWF-ATG GAA TCA AGA ATG GAA GGA GAT-3′; attB1-EFL3, 5′-GWF-ATG GAG GGA GAC ACA ATA TCT AGG-3′; attB1-EFL4, 5′-GWF-ATG GAA GGA GAT GTG TTG TCA GGA-3′; attB1-HvEU916968, 5′-GWF-ATG GAG AAC AGC AGC GGC CGG GAG-3′; attB1-InEU916969, 5′-GWF-ATG GAG AAC ACG TCA CGA GCC GTA-3′; attB1-PtEU916973, 5′-GWF-ATG GAG GGG GAA GCA TAT TCT GCT-3′; attB2-EFL1, 5′-GWR-AGA ACC GGT GGT GGT AGT TGT GGT G-3′; attB2-EFL2, 5′-GWR-CCC GGA TCT AAA TCT CTT CTG G-3′; attB2-EFL3, 5′-GWR-ATT AAG CAG GCC TGA TTC TTC T-3′; attB2-EFL4, 5′-GWR-ACC GGA TCT AAA TCT CTT CTG G-3′; attB2-ELF4, 5′-GWR-AGC TCT AGT TCC GGC AGC ACC A-3′; attB2-HvEU916968, 5′-GWR-CTG TGT GGG GCG CGG CCT CTT-3′; attB2-InEU916969, 5′-GWR-CTG TTC CCG GCC TCC GGA GCG-3′; and attB2-PtEU916973, 5′-GWR-AAA CTG AGG TCG AAA TCT CTT-3′.

The amplified fragments were inserted into the Gateway pDONR207 vector and then shuttled into pJawohl/ELF4p to complete construction design. The resultant plasmids were then introduced to *Agrobacterium tumefaciens* GV3101 pMPORK. The cloned pJalee4/*EFL* constructs were confirmed by sequencing and transformed into *elf4-1* plants harboring *CCA1*:*LUC* by a modified floral-dip method (Davis *et al.*, 2009). Three independent T2 lines for each T-DNA, which behaved similarly in at least two experiments, were included in this report.

Structure prediction

Automated ROSETTA predictions were exactly as described (Kolmos *et al.*, 2008). Advanced protein structure prediction, including secondary structure, disorder, and tertiary fold, was performed using GeneSilico metaserver (https://genesilico.pl/meta2) (Kurowski and Bujnicki, 2003). Prediction of potential coiled-coil regions was performed using several tools including COILS, MULTICOIL, MARCOIL, and PAIRCOIL2 (Delorenzi and Speed, 2002; Lupas *et al.*, 1991; McDonnell *et al.*, 2006; Wolf *et al.*, 1997). The three-dimensional structure modeling of the ELF4 monomer was carried out using a local installation of ROSETTA (Simons *et al.*, 1997). We generated 50,000 decoys and clustered them with lower bound for size of the top cluster set on 25, minimal clustering threshold 2, and maximal clustering threshold 5. A lowest-energy conformation from the largest cluster was considered as the most likely structure.

Modeling of the dimeric structure of ELF4 was done for the helical region of the monomer core with GRAMM, DOT, and ZDOCK with default parameters (Chen and Weng, 2002; Mandell *et al.*, 2001; Vakser, 1996). The 30 best-scoring decoys from each method were collected and clustered using MAXCLUSTER (www.sbg.bio.ic.ac.uk/~maxcluster/index.html). Dimer optimization was done with ROSETTADOCK (Schueler-Furman *et al.*, 2005), starting from the orientation corresponding to the largest cluster. The quality of models was assessed using PROQ and METAMQAP (Pawlowski *et al.*, 2008; Wallner and Elofsson, 2003).

ELF4 protein

The ELF4 coding region was inserted into pET-28a+vector (Novagen) using *Nde*I and *Xho*I cloning sites. Recombinant N-terminally His-tagged ELF4 was produced in BL21 (DE3)

E. coli strain (Novagen) and purified using HIS-Select nickel affinity gel (Sigma). The native-PAGE analysis was performed using 15% acidic native-polyacrylamide gel. Bands were photographed after Coomassie blue staining. The calibration standards used in the native-PAGE and gel filtration analysis were as follows: albumin, 67 kDa; chymotrypsinogen A, 25 kDa; ovalbumin, 43 kDa; and ribonuclease A, 13.7 kDa. Molecular mass estimation over gel filtration was performed to confirm the molecular mass of the purified ELF4 protein. For this, a Superdex 200 PC 3.2/30 (Amersham Biosciences) column was used with the flow rate 50 μl/min. Buffer: 250 mM NaCl, 50 mM sodium-phosphate buffer pH 8, and 10 mM β-mercaptoethanol.

Circular dichroism

Far-UV circular dichroism measurements were obtained using a JASCO model J-715 (JASCO, Gross-Umstadt, Germany). All measurements were performed at room temperature in a 1 mm path-length cuvette. For each measurement, 20 spectra were accumulated at a scan speed of 50 nm/min with a step resolution of 0.1 nm. The spectra were corrected for a protein-free spectrum obtained under identical conditions. Noise reduction was applied according to the JASCO software. Protein samples were at 0.14 and 0.22 mg/ml in 5 mM potassium-phosphate buffer pH 7.5. The spectra were normalized and the average was calculated. Secondary structure prediction was carried out using SELCON3 (Sreerama *et al.*, 1999; Whitmore and Wallace, 2008).

Growth conditions and luciferase imaging

Seeds were surface sterilized and plated on 2.2 g/l Murashige and Skoog (1× MS) medium (4.4 g/l; adjusted pH to 5.7) with 3% sucrose and 1% agar before being stratified at 4°C for 2–3 days and transferred to 12L:12D growth cabinets (100 μE white light). For both imaging experiments and RNA time courses, 7 day-old seedlings were used (see below). Luciferase imaging was performed using a Packard/PerkinElmer TopCount scintillation counter, as previously described (Hanano *et al.*, 2006). In general, a minimum of 24 seedlings per genotype was analyzed per experiment.

Rhythm analysis

Luminescence levels were quantified and analyzed essentially as described (McWatters *et al.*, 2000; Thain *et al.*, 2000), using the macrosuites TopTempII and Biological Rhythms Analysis Software System (BRASS) (Southern and Millar, 2005) (both available at http://millar.bio.ed.ac.uk), and fast Fourier transform nonlinear least squares (Plautz *et al.*, 1997). Sustainability (precision) of rhythms was derived from measurements of the RAE (Allen *et al.*, 2006; Izumo *et al.*, 2006).

RNA profiling

Total RNA was isolated with RNeasy Plant Mini Kit (Qiagen) from replicate samples of 1-week-old seedlings entrained in 12L:12D or those that were subsequently released into LL or DD. Total RNA (2 μg) was treated with DNAse I (Fermentas) and reverse-transcribed using oligo(dT)15 primer (Roche) and SuperScript II Reverse Transcriptase (Invitrogen) according to the manufacturers' instructions. qPCR was performed on 2 μl 1:10 diluted cDNA in a 20 μl reaction with IQ SYBR Green Supermix (Bio-Rad). All qPCRs were performed in the iCycler iQ5 Multicolor Real-time PCR Detection System (Bio-Rad). Each run included a standard curve, which was generated from serial dilutions of a pool of cDNA samples and a melting curve, which ensured amplification of one specific gene product. The amount of template in unknown samples was calculated from the threshold value by the iCycler software using the standard curve results. Measured transcript levels were normalized to the reference gene (*TUB2*) and subsequently also to the average diurnal level for the respective gene in wild type. The presented results are representative of a two entirely independent biological replicates, both of which were separately assessed in triplicate. Error bars represent adjusted standard deviation, calculated based on the pooled standard deviation

$$s_p = [(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \cdots (n_k - 1)s_k^2]/ [n_1 + n_2 + \cdots n_k - k]^{1/2}$$

and the propagation of error, $f = aA$; $\sigma_f^2 = a^2\sigma_A^2$.

CCA1, *TOC1*, and *TUB2* qPCR primers have been described (Ding *et al.*, 2007). Additional qPCR primers were LHY-fwd, 5′-GCT AAG GCA AGA AAG CCA TA-3′; LHY-rev, 5′-TGC CAA GCT CTT CCA TAA AG-3′; LUX-fwd, 5′-GCT TCG GAT AAG CTC TTC TCT TC-3′; LUX-rev, 5′-ATA AAC TGG CAT CTG CAT CAT CT-3′; GI-fwd, 5′-CTG TCT TTC TCC GTT GTT TCA CTG T-3′; GI-rev, 5′-TCA TTC CGT TCT TCT CTG TTG TTG G-3′; PRR7-fwd, 5′-TCT GAA GAG CTA ATG CAC GTG G-3′; PRR7-rev, 5′-ACA TGT GAG CGA TGA TTA TGG G-3′; PRR9-fwd, 5′-CAT CAA AAG CTT AGC CTC TCT G-3′; PRR9-rev, 5′-CTG TGG ACT GAA GAA CTT GGT TAC-3′; EFL1-fwd, 5′-GTA AAA TAA TGG AAG CAT CGA GAA A-3′; and EFL1-rev, 5′-ATC ACG ATT CTG ATC AAG ATA AAG C-3′.

Circadian modeling

The numerical simulations of *PRR9-ox* and *Y-ox* were performed using the three-loop model (Locke *et al.*, 2006) in the circadian modeling interface (available at http://millar.bio.ed.ac.uk). In this model *LHY* represents both *CCA1* and *LHY*, and *PRR9* is both *PRR9* and *PRR7*. The parameters were the initial level of mRNA, 14.69 nM; constitutive expression level, 8.07 nM/h; and constitutive translational rate constant, 0.29 1/h; for the *PRR9* constitutive gene expression. The parameters for the *Y* constitutive gene ex-

pression were initial level of mRNA, 0.0603 nM; constitutive expression level, 0.3 nM/h; and constitutive translational rate constant, 0.2485 1/h.

ACKNOWLEDGMENTS

We thank the Arabidopsis TILLING facility and NASC for seeds (TILLING and SALK lines), INRA for FLAG lines, and Mark R. Doyle (UW-Madison) for two *efl* lines. We acknowledge the following primary investigators for EST clones: H. Chaimovich, D.W. Choi, G. Chuck, T. Close, R. Croteau, R. Dean, S. Iida, W.R. McCombie, R.W. Michelmore, L. Pratt, Z. Schwarz-Sommer, S. Stanfield, S. Tanksley, M.K. Udvardi, C.P. Vance, B. Weisshaar, T.A. Wilkins, and R. Wise. E.K. thanks A.M. Davis for technical assistance and A. Giakountis for the *GI* primer. We are grateful to L. Kozma-Bognar for comments on the paper. E.K., H.S, and S.J.D. received funding from the Max-Planck-Gesellschaft, and for E.K. and S.J.D., the Deutsche Forschungsgemeinschaft (Grant No. DA 1061/4-1); F.N. is supported by a Howard Hughes Medical Institute International Fellowship. M.N., M.W., and J.M.B. were supported by the National Institutes of Health (Grant No. 1R01GM081680-01).

REFERENCES

Alabadi, D, Oyama, T, Yanovsky, MJ, Harmon, FG, Mas, P, and Kay, SA (2001). "Reciprocal regulation between *TOC1* and *LHY/CCA1* within the Arabidopsis circadian clock." *Science* **293**, 880–883.

Alabadi, D, Yanovsky, MJ, Mas, P, Harmer, SL, and Kay, SA (2002). "Critical role for CCA1 and LHY in maintaining circadian rhythmicity in Arabidopsis." *Curr. Biol.* **12**, 757–761.

Allen, T, Koustenis, A, Theodorou, G, Somers, DE, Kay, SA, Whitelam, GC, and Devlin, PF (2006). "Arabidopsis *FHY3* specifically gates phytochrome signaling to the circadian clock." *Plant Cell* **18**, 2506–2516.

Altschul, SF, Gish, W, Miller, W, Myers, EW, and Lipman, DJ (1990). "Basic local alignment search tool." *J. Mol. Biol.* **215**, 403–410.

Bao, X, Franks, RG, Levin, JZ, and Liu, Z (2004). "Repression of *AGAMOUS* by *BELLRINGER* in floral and inflorescence meristems." *Plant Cell* **16**, 1478–1489.

Boxall, SF, Foster, JM, Bohnert, HJ, Cushman, JC, Nimmo, HG, and Hartwell, J (2005). "Conservation and divergence of circadian clock operation in a stress-inducible Crassulacean acid metabolism species reveals clock compensation against stress." *Plant Physiol.* **137**, 969–982.

Chen, R, and Weng, Z (2002). "Docking unbound proteins using shape complementarity, desolvation, and electrostatics." *Proteins* **47**, 281–294.

Davis, AM, Hall, A, Millar, AJ, Darrah, C, and Davis, SJ (2009). "Protocol: streamlined sub-protocols for floral-dip transformation and selection of transformants in *Arabidopsis thaliana*." *Plant Methods* **5**, 3.

Delorenzi, M, and Speed, T (2002). "An HMM model for coiled-coil domains and a comparison with PSSM-based predictions." *Bioinformatics* **18**, 617–625.

Ding, Z, Doyle, MR, Amasino, RM, and Davis, SJ (2007). "A complex genetic interaction between *Arabidopsis thaliana TOC1* and *CCA1*/*LHY* in driving the circadian clock and in output regulation." *Genetics* **176**, 1501–1510.

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

Dowson-Day, MJ, and Millar, AJ (1999). "Circadian dysfunction causes aberrant hypocotyl elongation patterns in Arabidopsis." *Plant J.* **17**, 63–71.

Doyle, MR (2003). "The cloning and characterization of EARLY FLOWERING4: a gene involved in circadian regulation and the control of flowering time in Arabidopsis thaliana." Ph.D. dissertation, University of Wisconsin-Madison.

Doyle, MR, Davis, SJ, Bastow, RM, McWatters, HG, Kozma-Bognar, L, Nagy, F, Millar, AJ, and Amasino, RM (2002). "The ELF4 gene controls circadian rhythms and flowering time in *Arabidopsis thaliana*." *Nature (London)* **419**, 74–77.

Enns, LC, Kanaoka, MM, Torii, KU, Comai, L, Okada, K, and Cleland, RE (2005). "Two callose synthases, GSL1 and GSL5, play an essential and redundant role in plant and pollen development and in fertility." *Plant Mol. Biol.* **58**, 333–349.

See EPAPS Document No. E-HJFOA5-3-004906 for supplemental material. This document can be reached through a direct link in the online article's HTML reference section or via the EPAPS homepage (http://www.aip.org/pubservs/epaps.html).

Farre, EM, Harmer, SL, Harmon, FG, Yanovsky, MJ, and Kay, SA (2005). "Overlapping and distinct roles of *PRR7* and *PRR9* in the Arabidopsis circadian clock." *Curr. Biol.* **15**, 47–54.

Fowler, S, Lee, K, Onouchi, H, Samach, A, Richardson, K, Morris, B, Coupland, G, and Putterill, J (1999). "*GIGANTEA*: a circadian clock-controlled gene that regulates photoperiodic flowering in Arabidopsis and encodes a protein with several possible membrane-spanning domains." *EMBO J.* **18**, 4679–4688.

Fukushima, A, Kusano, M, Nakamichi, N, Kobayashi, M, Hayashi, N, Sakakibara, H, Mizuno, T, and Saito, K (2009). "Impact of clock-associated Arabidopsis pseudo-response regulators in metabolic coordination." *Proc. Natl. Acad. Sci. U.S.A.* **106**, 7251–7256.

Hanano, S, Domagalska, MA, Nagy, F, and Davis, SJ (2006). "Multiple phytohormones influence distinct parameters of the plant circadian clock." *Genes Cells* **11**, 1381–1392.

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

Hazen, SP, Schultz, TF, Pruneda-Paz, JL, Borevitz, JO, Ecker, JR, and Kay, SA (2005). "*LUX ARRHYTHMO* encodes a Myb domain protein essential for circadian rhythms." *Proc. Natl. Acad. Sci. U.S.A.* **102**, 10387–10392.

Herzog, ED (2007). "Neurons and networks in daily rhythms." *Nat. Rev. Neurosci.* **8**, 790–802.

Hicks, KA, Millar, AJ, Carre, IA, Somers, DE, Straume, M, Meeks-Wagner, DR, and Kay, SA (1996). "Conditional circadian dysfunction of the Arabidopsis *early-flowering 3* mutant." *Science* **274**, 790–792.

Huelsenbeck, JP, and Ronquist, F (2001). "MRBAYES: Bayesian inference of phylogenetic trees." *Bioinformatics* **17**, 754–755.

Ito, S, Nakamichi, N, Nakamura, Y, Niwa, Y, Kato, T, Murakami, M, Kita, M, Mizoguchi, T, Niinuma, K, Yamashino, T, and Mizuno, T (2007). "Genetic linkages between circadian clock-associated components and phytochrome-dependent red light signal transduction in *Arabidopsis thaliana*." *Plant Cell Physiol.* **48**, 971–983.

Ito, S, Niwa, Y, Nakamichi, N, Kawamura, H, Yamashino, T, and Mizuno, T (2008). "Insight into missing genetic links between two evening-expressed pseudo-response regulator genes *TOC1* and *PRR5* in the circadian clock-controlled circuitry in *Arabidopsis thaliana*." *Plant Cell Physiol.* **49**, 201–213.

Izumo, M, Sato, TR, Straume, M, and Johnson, C (2006). "Quantitative analyses of circadian gene expression in mammalian cell cultures." *PLOS Comput. Biol.* **2**, e136.

Khanna, R, Kikis, EA, and Quail, PH (2003). "*EARLY FLOWERING 4* functions in phytochrome B-regulated seedling de-etiolation." *Plant Physiol.* **133**, 1530–1538.

Kikis, EA, Khanna, R, and Quail, PH (2005). "*ELF4* is a phytochrome-regulated component of a negative-feedback loop involving the central oscillator components *CCA1* and *LHY*." *Plant J.* **44**, 300–313.

Kolmos, E, and Davis, SJ (2007). "*ELF4* as a central gene in the circadian clock." *Plant Signal. Behav.* **2**, 370–372.1559-2316

Kolmos, E, Schoof, H, Plumer, M, and Davis, SJ (2008). "Structural insights into the function of the core-circadian factor TIMING OF CAB2 EXPRESSION 1 (TOC1)." *J. Circadian Rhythms* **6**, 3.

Krysan, PJ, Young, JC, and Sussman, MR (1999). "T-DNA as an insertional mutagen in Arabidopsis." *Plant Cell* **11**, 2283–2290.
Kurowski, MA, and Bujnicki, JM (2003). "GeneSilico protein structure prediction meta-server." *Nucleic Acids Res.* **31**, 3305–3307.
Lewis, EB (1941). "Another case of unequal crossing-over in *Drosophila melanogaster*." *Proc. Natl. Acad. Sci. U.S.A.* **27**, 31–34.
Locke, JCW, Kozma-Bognar, L, Gould, PD, Feher, B, Kevei, E, Nagy, F, Turner, MS, Hall, A, and Millar, AJ (2006). "Experimental validation of a predicted feedback loop in the multi-oscillator clock of *Arabidopsis thaliana*." *Mol. Syst. Biol.* **2**, 59.
Lupas, A, Van Dyke, M, and Stock, J (1991). "Predicting coiled coils from protein sequences." *Science* **252**, 1162–1164.
Mandell, JG, Roberts, VA, Pique, ME, Kotlovyi, V, Mitchell, JC, Nelson, E, Tsigelny, I, and Ten Eyck, LF (2001). "Protein docking using continuum electrostatics and geometric fit." *Protein Eng.* **14**, 105–113.
Martin-Tryon, EL, Kreps, JA, and Harmer, SL (2007). "*GIGANTEA* acts in blue light signaling and has biochemically separable roles in circadian clock and flowering time regulation." *Plant Physiol.* **143**, 473–486.
McDonnell, AV, Jiang, T, Keating, AE, and Berger, B (2006). "Paircoil2: improved prediction of coiled coils from sequence." *Bioinformatics* **22**, 356–358.
McWatters, HG, Bastow, RM, Hall, A, and Millar, AJ (2000). "The *ELF3* zeitnehmer regulates light signalling to the circadian clock." *Nature (London)* **408**, 716–720.
McWatters, HG, Kolmos, E, Hall, A, Doyle, MR, Amasino, RM, Gyula, P, Nagy, F, Millar, AJ, and Davis, SJ (2007). "*ELF4* is required for oscillatory properties of the circadian clock." *Plant Physiol.* **144**, 391–401.
Meinke, DK, and Koornneef, M (1997). "Community standards for Arabidopsis genetics." *Plant J.* **12**, 247–253.
Millar, AJ (2003). "A suite of photoreceptors entrains the plant circadian clock." *J. Biol. Rhythms* **18**, 217–226.
Mizoi, J, Nakamura, M, and Nishida, I (2006). "Defects in *CTP:PHOSPHORYLETHANOLAMINE CYTIDYLYLTRANSFERASE* affect embryonic and postembryonic development in Arabidopsis." *Plant Cell* **18**, 3370–3385.
Murakami, M, Tago, Y, Yamashino, T, and Mizuno, T (2007). "Comparative overviews of clock-associated genes of *Arabidopsis thaliana* and *Oryza sativa*." *Plant Cell Physiol.* **48**, 110–121.
Nakamichi, N, Kita, M, Ito, S, Sato, E, Yamashino, T, and Mizuno, T (2005a). "The Arabidopsis pseudo-response regulators, *PRR5* and *PRR7*, coordinately play essential roles for circadian clock function." *Plant Cell Physiol.* **46**, 609–619.
Nakamichi, N, Kita, M, Ito, S, Yamashino, T, and Mizuno, T (2005b). "*PSEUDO-RESPONSE REGULATORS, PRR9, PRR7* and *PRR5*, together play essential roles close to the circadian clock of *Arabidopsis thaliana*." *Plant Cell Physiol.* **46**, 686–698.
Nozue, K, Covington, MF, Duek, PD, Lorrain, S, Fankhauser, C, Harmer, SL, and Maloof, JN (2007). "Rhythmic growth explained by coincidence between internal and external cues." *Nature (London)* **448**, 358–361.
Onai, K, and Ishiura, M (2005). "*PHYTOCLOCK 1* encoding a novel GARP protein essential for the Arabidopsis circadian clock." *Genes Cells* **10**, 963–972.
Pawlowski, M, Gajda, MJ, and Matlak, R (2008). "MetaMQAP: a meta-server for the quality assessment of protein models." *BMC Bioinf.* **9**, 403.
Perales, M, and Mas, P (2007). "A functional link between rhythmic changes in chromatin structure and the Arabidopsis biological clock." *Plant Cell* **19**, 2111–2123.
Plautz, JD, Straume, M, Stanewsky, R, Jamison, CF, Brandes, C, Dowse, HB, Hall, JC, and Kay, SA (1997). "Quantitative analysis of *Drosophila period* gene transcription in living animals." *J. Biol. Rhythms* **12**, 204–217.
Pruneda-Paz, JL, Breton, G, Para, A, and Kay, SA (2009). "A functional genomics approach reveals CHE as a component of the Arabidopsis circadian clock." *Science* **323**, 1481–1485.
Ronquist, F, and Huelsenbeck, JP (2003). "MrBayes 3: Bayesian phylogenetic inference under mixed models." *Bioinformatics* **19**, 1572–1574.
Sali, A, and Blundell, TL (1993). "Comparative protein modelling by satisfaction of spatial restraints." *J. Mol. Biol.* **234**, 779–815.
Salome, PA, Xie, Q, and McClung, CR (2008). "Circadian timekeeping during early Arabidopsis development." *Plant Physiol.* **147**, 1110–1125.
Schaffer, R, Ramsay, N, Samach, A, Corden, S, Putterill, J, Carre, IA, and Coupland, G (1998). "The *late elongated hypocotyl* mutation of Arabidopsis disrupts circadian rhythms and the photoperiodic control of flowering." *Cell* **93**, 1219–1229.
Schueler-Furman, O, Wang, C, and Baker, D (2005). "Progress in protein-protein docking: atomic resolution predictions in the CAPRI experiment using RosettaDock with an improved treatment of side-chain flexibility." *Proteins* **60**, 187–194.
Simons, KT, Kooperberg, C, Huang, E, and Baker, D (1997). "Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions." *J. Mol. Biol.* **268**, 209–225.
Southern, MM, and Millar, AJ (2005). "Circadian genetics in the model higher plant, *Arabidopsis thaliana*." *Methods Enzymol.* **393**, 23–35.
Sreerama, N, Venyaminov, SY, and Woody, RW (1999). "Estimation of the number of alpha-helical and beta-strand segments in proteins using circular dichroism spectroscopy." *Protein Sci.* **8**, 370–380.
Strayer, C, Oyama, T, Schultz, TF, Raman, R, Somers, DE, Mas, P, Panda, S, Kreps, JA, and Kay, SA (2000). "Cloning of the Arabidopsis clock gene *TOC1*, an autoregulatory response regulator homolog." *Science* **289**, 768–771.
Swofford, DL, 2002. *Phylogenetic Analysis Using Parsimony (*and Other Methods), Ver. 4.0b10*, Sinauer, Sunderland, MA.
Thain, SC, Hall, A, and Millar, AJ (2000). "Functional independence of circadian clocks that regulate giant gene expression." *Curr. Biol.* **10**, 951–956.
Thompson, JD, Higgins, DG, and Gibson, TJ (1994). "CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice." *Nucleic Acids Res.* **22**, 4673–4680.
Vakser, IA (1996). "Low-resolution docking: prediction of complexes for underdetermined structures." *Biopolymers* **39**, 455–464.
Wallner, B, and Elofsson, A (2003). "Can correct protein models be identified?" *Protein Sci.* **12**, 1073–1086.
Wang, ZY, and Tobin, EM (1998). "Constitutive expression of the *CIRCADIAN CLOCK ASSOCIATED 1* (*CCA1*) gene disrupts circadian rhythms and suppresses its own expression." *Cell* **93**, 1207–1217.
Whitmore, L, and Wallace, BA (2008). "Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases." *Biopolymers* **89**, 392–400.
Wolf, E, Kim, PS, and Berger, B (1997). "MultiCoil: a program for predicting two- and three-stranded coiled coils." *Protein Sci.* **6**, 1179–1189.
Zeilinger, MN, Farre, EM, Taylor, SR, Kay, SA, and Doyle, FJ (2006). "A novel computational model of the circadian clock in Arabidopsis that incorporates *PRR7* and *PRR9*." *Mol. Syst. Biol.* **2**, 58.
Zwickl, DJ (2006). "Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion." Ph.D. dissertation, The University of Texas at Austin.