# Supporting Information Appendix for

# Correct biological timing in *Arabidopsis* requires multiple light signaling pathways

N Dalchau, KE Hubbard, FC Robertson, CT Hotta, HM Briggs, G-B Stan, JM Gonçalves, AAR Webb

## SI Text

### Linear Time-Invariant (LTI) Systems with Delayed Inputs

We propose that the regulation of rhythmic biological processes by the circadian signaling network may be described by linear time-invariant (LTI) systems with delayed inputs. Linear systems have the desirable property of maintaining the period of oscillatory inputs. Therefore, circadian regulation by LTI systems guarantees circadian rhythms in the output process. Delayed inputs correspond to the uncharacterized pathways connecting the input variables, which are measurements of *CCA1* expression (a proxy for the circadian central oscillator) and the availability of light. The output signal $y$ represents measured aequorin (AEQ) luminescence data. A state-space representation of an LTI system with delayed inputs $u_\tau(t)$ and output $y(t)$ is written as

$$\frac{dx}{dt}(t) = Ax(t) + Bu_\tau(t), \quad (1)$$

$$y(t) = Cx(t) + d \quad (2)$$

where $x \in \mathbb{R}^n$ is the vector of $n$ state variables, $u_\tau(t) \in \mathbb{R}^2$ is the delayed input vector*, $d \in \mathbb{R}$ is a constant offset of the measured output $y(t)$, $A \in \mathbb{R}^{n\times n}$ is the matrix of interconnections, $B \in \mathbb{R}^{n\times 2}$ is the input matrix and $C \in \mathbb{R}^{1\times n}$ is the output vector. Since $y$ is the measurement of a single state, $C$ consists of a 1 in the first entry and 0's in the entries corresponding to the $n-1$ hidden states. The time-invariant property of this model class refers to the system parameters $A$, $B$, $C$ and $d$ (and delays $\tau_{CCA1}$, $\tau_{\text{light}}$) being constant.

*Remark* 1. We assumed that regulation by clock-sensitive and light/dark-sensitive pathways is independent. The assumption is reasonable provided that we carefully distinguish between light/dark regulation mediated by the circadian clock and light/dark regulation not mediated by the clock. While the availability of light is trivially independent of the clock, the signal transduction pathway may not be. Therefore, if $\tau_{\text{light}} > 0$, then the independence assumption may not be appropriate.

*Remark* 2. Systems with delayed inputs may also be thought of as infinite-dimensional systems. Suppose $X(s)$ is the Laplace transform of $x(t)$, $U(s)$ is the Laplace transform of $u(t)$ etc., then Eq. 1 may be written in the Laplace domain (after rearrangement) as

$$X(s) = G(s)e^{-s\tau}U(s), \quad (3)$$

where $G(s) = (sI - A)^{-1}B$ is the transfer function corresponding to a standard linear system (with no delays and full state observability, i.e., $C = I$). As the exponential term is a polynomial of infinite degree, it may be represented as an infinite-dimensional linear system, or approximated by a finite-dimensional system.

*Remark* 3. The provision of hidden variables ($n > 1$) leads to degrees of freedom in the system matrices $A$ and $B$. This is due to the input-output invariance of transfer functions with respect to similarity transformations: an arbitrary linear system $\dot{x} = Ax + Bu$, $y = Cx$ has the transfer function representation $Y(s) = C(sI - A)^{-1}BU(s)$. Applying a similarity transformation $T \in \mathbb{R}^{n\times n}$ such that $x = Tz$ gives

$$\dot{x} = T\dot{z} = ATz + Bu$$

Taking Laplace transforms and noting that $Y(s) = CTZ(s)$ gives

$$\begin{aligned} Y(s) &= CT(sI - T^{-1}AT)^{-1}T^{-1}BU(s) \\ &= C(sI - A)^{-1}BU(s), \end{aligned}$$

as before. The degrees of freedom resulting from this invariance to similarity transformation may be fixed (for convenience) by writing the LTI system in the controller canonical form

$$\frac{dx}{dt} = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & \dots & 0 & 1 \\ a_1 & a_2 & a_3 & \dots & a_n \end{pmatrix} x + \begin{pmatrix} b_{11} & \dots & b_{1m} \\ \vdots & & \vdots \\ b_{n1} & \dots & b_{nm} \end{pmatrix} u, \quad (4)$$

where each $a_i, b_{ij} \in \mathbb{R}$.

*The notation $u_\tau(t)$ is shorthand for a vector of inputs $u_1, u_2, \dots$ subject to distinct delays $\tau_1, \tau_2, \dots$. Since we consider two inputs, this amounts here to $u_\tau(t) = \begin{bmatrix} u_{CCA1}(t - \tau_{CCA1}) & u_{\text{light}}(t - \tau_{\text{light}}) \end{bmatrix}^T$.

## Model Identification

Systems Identification theory provides powerful tools for the estimation of LTI systems from input-output data alone, enabling also the provision of hidden variables. However, there have been no published algorithms which approximate the delays $\tau$ in addition to the system matrices $A$ and $B$ in Eq. 1. Instead of estimating $\tau$ explicitly, we pre-processed the input-output data to account for different choices of $\tau$, conducted parameter estimation for each choice, and compared the performance of the obtained models[†]. This essentially reduces the estimation problem to the basic no-delay case. Therefore, model identification was a two-stage process, consisting of data preparation and parameter estimation.

### Parameter Estimation

The prediction-error method (PEM) (1) was used to estimate the system matrices $A$ and $B$ and the constant offset $d$ for each choice of $\tau$ and $n$. PEM falls under the class of indirect methods for systems identification, whereby a discretized version of an LTI system is first computed, before transforming it to a continuous-time representation. PEM requires that the input-output observations are uniformly spaced in time and assumes a zero-order hold (ZOH) intersample behavior. Estimation of $d$ required a re-parameterization of the LTI model class, as a constant offset in the output signal cannot be treated by PEM. If we define $x_d := \begin{pmatrix} x \\ d \end{pmatrix}$, then equations 1 & 2 may be written as

$$\begin{aligned} \dot{x}_d &= \begin{pmatrix} A & 0 \\ 0 & 0 \end{pmatrix} x_d + \begin{pmatrix} B \\ 0 \end{pmatrix} u \\ y &= \begin{pmatrix} C & 1 \end{pmatrix}, \end{aligned} \tag{5}$$

providing a structured re-parameterization whereby $d$ can be estimated as the initial condition for the last entry in $x_d$.

## Cytosolic-free $Ca^{2+}$

In this section, we describe the derivation of an input-delayed linear system model for the regulation of cytosolic-free $Ca^{2+}$ ($[Ca^{2+}]_{cyt}$). Our experimental observations of $[Ca^{2+}]_{cyt}$ exist in the form of bioluminescence measurements of AEQ, a $Ca^{2+}$-sensitive bioluminescent protein which has been constitutively expressed in *Arabidopsis*. The derivation procedure outlined here was also applied to models of *CAB2::luc* expression and rhythmically expressed transcripts.

### Pre-Processing the Input-Output Data

The target model output during model estimation was taken to be the measured AEQ luminescence values obtained from 10 day old seedlings measured for 2 days of 12L/12D cycles followed by 3 days of DD (Fig. S1B). Normalized measurements of *CCA1::luc* luminescence in the same conditions were used as the values for $u_{CCA1}$. The values were normalized by dividing all measurement values by the maximum value in 12L/12D cycles (Fig. S1A). Normalization of the *CCA1::luc* luminescence data was required to enable model simulation with contrasting datasets (alternative photoperiodic conditions) which may have different luminescent signal strengths. We were prevented from using current literature mathematical models of the *Arabidopsis* circadian clock for the *CCA1* input, as these models correspond to plants grown on agar media containing 3% (w/v) sucrose (2–5). Our plants were grown in the absence of exogenous sucrose to permit circadian oscillations of $[Ca^{2+}]_{cyt}$ (6). $u_{light}$ was set equal to 1 during light and 0 during dark.

The sets of input-output triples ($u_{CCA1}$, $u_{light}$, AEQ) were formed for each combination of ($\tau_{CCA1}$, $\tau_{light}$) with $\tau_{CCA1}$ = 0, 0.1, 0.2, . . ., 10 h and $\tau_{light}$ = 0, 1, 2, 3, 4, 5 h. The values of $u_{light}$ were set as a binary sequence of 1's and 0's corresponding to the experimental conditions of the output estimation dataset, whilst reflecting the value of $\tau_{light}$ being used (Fig. S1A). As PEM uses zero-order hold intersample behavior, the assignment of $u_{light}$ for estimation was exact. Since *CCA1* expression is a continuous variable (which correlates with the cell-averaged concentration of *CCA1* mRNA), it was more natural to use a first-order hold intersample behavior (linear interpolation). As AEQ luminescence was measured at t = 1, 3, 5, . . . and PEM requires ZOH, it was not possible to assign a change in the value of $u_{light}$ at light-dark or dark-light transitions unless $\tau_{light}$ = 1, 3, 5, . . . also. To enable estimation of models with $\tau_{light}$ being allowed to take any integer value, the mid-points lying between successive values of AEQ luminescence were computed by linear interpolation (AEQ is a continuous variable which relates to the cell-averaged $[Ca^{2+}]_{cyt}$), and the obtained interpolated values were appended to the estimation dataset.

### Model Selection

As the number of derived models was large, with many combinations of $\tau_{light}$, $\tau_{CCA1}$ and $n$ considered, stringent selection criteria were developed. Furthermore, for models with hidden variables (i.e., $n > 1$), it was possible to describe the target output data with arbitrary closeness. As such, it was important to seek a compromise between the complexity of the model and the reproducibility of the target data. Each estimated model was analyzed by assessing the predictability of all models with respect to alternative datasets, or 'cross-validation'.

A cross-validation procedure was applied to evaluate the wider applicability of the derived LTI models and to therefore establish the most suitable choices of $n$, $\tau_{CCA1}$ and

[†]Pre-specification of the delay vector $\tau$ is equivalent to the required pre-specification of the model order $n$. Appropriate values may be selected after analyzing model performance.

$\tau_{\text{light}}$. Input-output datasets (AEQ and *CCA1::luc* luminescence) were collected in photoperiodic conditions which contrasted with those used for model estimation: cycles of 12L/12D before transfer to LL and cycles of long (16 h) and short (8 h) photoperiods (Supplementary Figs. S1C, S1E, and S1G). The ability of any given model to match the experimental data was investigated by computing Pearson's product-moment sample correlation coefficient between the measured data and the simulated output data at the corresponding time points. For any vectors $x \in \mathbb{R}^N$ and $y \in \mathbb{R}^N$, the sample correlation coefficient $r$ is defined as

$$r(x,y) := \sum_{i=1}^{N} \left( \frac{x_i - \bar{x}}{||x - \bar{x}\mathbf{1}||_2} \frac{y_i - \bar{y}}{||y - \bar{y}\mathbf{1}||_2} \right) \tag{6}$$

where $\bar{x} \in \mathbb{R}$ and $\bar{y} \in \mathbb{R}$ are the sample means of $x$ and $y$ respectively, $||x||_2$ is the euclidian norm of $x$, and $\mathbf{1} \in \mathbb{R}^N$ is the vector $(1, \ldots, 1)^T$. For each of the wild-type datasets available for estimation and cross-validation, $r_D$ (12L/12D–DD), $r_L$ (12L/12D–LL), $r_{LD}$ (16L/8D) and $r_{SD}$ (8L/16D) are defined accordingly to measure the correlation between measured and simulated outputs. To quantify the general model performance over each of the four wild-type input-output datasets available, we defined a weighted sample correlation $r_w$ as

$$r_w := n_D r_D + n_L r_L + n_{LD} r_{LD} + n_{SD} r_{SD} \tag{7}$$

where $n_k$ is the number of points in dataset $k \in \{D, L, LD, SD\}$.

2nd order models with $\tau_{CCA1}$ near 5 h offered the best predictive capability when comparing the cross-validation measures $r_w$ of the different models. An optimal $\tau_{CCA1}$ (i.e., a $\tau_{CCA1}$ which had the highest $r_w$) was chosen for each combination of $\tau_{\text{light}}$ and $n$ (Fig. S2). The difference in performance between 2nd and 3rd order models was very small (Fig. S2D), which was confirmed by comparing the simulated outputs in each of the 4 photoperiods investigated (Fig. S3). Therefore, $n = 2$ can be considered as a sufficient minimal description of wild-type dynamical behaviors.

### Isolation of Input Pathways

Through model estimation and cross-validation, it was possible to determine a minimal order ($n = 2$) and a narrow range of durations for the uncharacterized pathway connecting *CCA1* expression and $[Ca^{2+}]_{\text{cyt}}$ (4.8 h $\lesssim \tau_{CCA1} \lesssim$ 5.2 h; Fig. S2B), though it was not possible to distinguish the duration of the light/dark input pathway, $\tau_{\text{light}}$. The light/dark input pathway was isolated by simulating a mutation in *CCA1*, achieved by setting $u_{CCA1} = 0$ throughout, while leaving $u_{\text{light}}$ as for the wild-type case. In this way, the effect of different values of $\tau_{\text{light}}$ could be seen in any photoperiodic regime. The optimal-$\tau_{CCA1}$ 2nd order models were simulated in 12L/12D cycles and compared with the corresponding measurements of AEQ luminescence in *cca1*-1 mutants (Fig. S4). The timing of the increase in simulated AEQ luminescence in a *cca1*-1 mutant following a dark-light transition was correlated with the value of $\tau_{\text{light}}$ (Fig. S4E). Since experimentally observed $[Ca^{2+}]_{\text{cyt}}$ increased immediately following a dark-light transition in *cca1*-1 mutants (7) (Fig. S4E) , we reason that 0 h is the best choice for $\tau_{\text{light}}$. Therefore, the candidate model estimated with PEM (for $n = 2$, $\tau_{\text{light}} = 0$ h, $\tau_{CCA1} = 5.2$ h) is

$$\begin{pmatrix} \dot{x_1}(t) \\ X2(t) \end{pmatrix} = A \begin{pmatrix} x_1(t) \\ X2(t) \end{pmatrix} + B \begin{pmatrix} u_{CCA1}(t - 5.2) \\ u_{\text{light}}(t) \end{pmatrix} \tag{8}$$

$$\text{AEQ} = x_1 + 14087 \tag{9}$$

$$\text{where} \quad A = \begin{pmatrix} 0 & 1 \\ -0.0313 & -0.3252 \end{pmatrix}, \tag{10}$$

$$B = \begin{pmatrix} 2445.8 & 1083.4 \\ -1187.3 & -390.96 \end{pmatrix} \tag{11}$$

This latter candidate model was also simulated for a mutation in *CCA1* in the photoperiodic conditions used for cross-validation, and compared with the corresponding wild-type simulations (Fig. S4).

### Mutation in the Hidden State Variable

The LTI state-space model class used to describe the circadian and light/dark regulation of $[Ca^{2+}]_{\text{cyt}}$ has only a single output. Without loss of generality, we can chose the observable output to be one of the state variables, with the other $n - 1$ variables being *hidden*. The dynamics of hidden variables cannot be simulated, a consequence of the invariance of transfer functions to similarity transformations (explained above). To enable further investigation of our candidate 2nd order model, we constructed a generalized form of the system matrices $A$ and $B$ which parameterized the set of all equivalent state-space realizations.

Consider a similarity matrix $T$ applied to a 2nd order linear system $\dot{x} = Ax + Bu$ such that $x = T\hat{x}$. In terms of the arbitrary vector of state variables $\hat{x}$, the system evolves as

$$\dot{\hat{x}} = \hat{A}\hat{x} + \hat{B}u \tag{12}$$

where $\hat{A} = T^{-1}AT$ and $\hat{B} = T^{-1}B$ are the generalized system matrices. Suppose $T = \begin{pmatrix} 1 & 0 \\ t_1 & t_2 \end{pmatrix}$ $(t_1, t_2 \in \mathbb{R})$ so that the (measurable) first entry in the state vector is preserved (i.e., $\hat{x}_1 = x_1$), and $A$, $B$ are in the canonical form of Eq. 4. Then by a series of simple matrix computations, the generalized system matrices may be written as

$$\hat{A} = T^{-1}AT = \begin{pmatrix} t_1 & t_2 \\ \dfrac{-t_1^2 + a_2 t_1 + a_1}{t_2} & -t_2 + a_2 \end{pmatrix} \tag{13}$$

$$\hat{B} = T^{-1}B = \begin{pmatrix} b_{11} & b_{12} \\ \dfrac{-b_{11}t_1 + b_{21}}{t_2} & \dfrac{-b_{12}t_1 + b_{22}}{t_2} \end{pmatrix} \tag{14}$$

In this latter generalized system, the state vector is composed of a variable representing AEQ ($x_1$; AEQ shifted by $d$ luminescence units) and a variable $X2$ which is parameterized by the choice of $t_1$ and $t_2$. To simulate a mutation in $X2$, we simply set it equal to 0 throughout, thus reducing the model to a 1st order system

$$\dot{x}_1 = t_1 x_1 + \hat{b}_1 u, \tag{15}$$

where $\hat{b}_1$ is the first row of $\hat{B}$. The dynamics of a simulated *x2* mutant therefore depends only on the choice of $t_1$. Note further that $t_1$ also determines the stability of the system.

While it was not possible to determine a suitable value for $t_1$ (or $t_2$) from the currently available experimental data, it was possible to predict dynamical behaviors in system perturbations. A series of simulations of Eq. 15 were carried out to obtain predictions of $[Ca^{2+}]_{cyt}$ in an *x2* mutant. The system was simulated with $t_1 = -0.01, -0.1, -0.25, -0.5$ and -1, subject to the input signals that were used previously for cross-validation (Fig. S6). The most striking observations were: 1) if $t_1$ was not chosen to be sufficiently negative then unstable dynamics could be observed as unbounded increases or decreases in simulated AEQ, and 2) irrespective of the value of $t_1$, there was a delayed increase in simulated AEQ following a dark to light transition in 16L/8D (Fig. S6C).

The generalized LTI system (Eq. 12) was also used to investigate the interactions between AEQ, *X2*, *CCA1* and light/dark. The internal structure of LTI systems with hidden states is not fixed, as similarity transformations can re-structure the state variables while preserving the input-output behavior. Using the second-order generalized system to establish whether particular entries in $A$ and $B$ could be zero, we found that no structural information could be derived from our model. From Eq. 13, the (2,1) element of $\hat{A}$ is equal to zero provided‡

$$t_1 = \frac{a_2}{2} \pm \frac{\sqrt{a_2^2 + 4a_1}}{2}, \tag{16}$$

which is only possible if $a_2^2 + 4a_1 \geq 0$. Using the parameter values of the candidate model (Eq. 8), we find that $a_2^2 + 4a_1 = -0.0195$, implying that it is not possible to choose $t_1$ such that $\hat{a}_{21}$ is zero. Therefore, *X2* is necessarily regulated by $[Ca^{2+}]_{cyt}$. The (2,1) entry of $\hat{B}$ is zero if and only if $t_1 = \frac{b_{21}}{b_{11}} = -0.4855$, while the (2,2) entry is zero if and only if $t_1 = \frac{b_{22}}{b_{12}} = -0.3609$. As it is possible to assign $t_1$ such that specific entries of $B$ are zero, it is not possible to determine properties of the input structure.

**Bode Frequency Response Analysis**

One of the main characteristics of LTI systems is that their response to sinusoidal signals is a sinusoidal signal with the same frequency.§ Only the amplitude of the output with respect to the amplitude of the input (the *gain*) or the phase of the output with respect to the phase of a periodic input may change. Therefore, LTI systems can be characterized through their frequency response, which is best visualized using Bode magnitude and phase plots.

Bode magnitude and phase analysis was conducted for the pathways connecting *CCA1* to $[Ca^{2+}]_{cyt}$ and light/dark to $[Ca^{2+}]_{cyt}$ (Fig. 3A). The frequency-domain transfer function representation of the candidate model (Eq. 8) was obtained by taking the Laplace transform, which yields

$$\mathcal{L}\{\mathrm{AEQ}(t)\} = \begin{pmatrix} G_{CCA1}(s) \\ G_{\mathrm{light}}(s) \end{pmatrix} \mathcal{L}\{ \begin{pmatrix} u_{CCA1}(t-5.2) \\ u_{\mathrm{light}}(t) \end{pmatrix} \}, \tag{17}$$

where

$$G_{CCA1}(s) = \frac{2445.8s - 391.9}{s^2 + 0.3252s + 0.0313}, \tag{18}$$

$$G_{\mathrm{light}}(s) = \frac{1083.4s - 38.6}{s^2 + 0.3252s + 0.0313} \tag{19}$$

From the frequency-domain representation, the magnitude of the frequency response represents the input-output gain at each frequency and is obtained in dB units as $20 \log |G(j\omega)|$, where $\omega$ is the input frequency. The phase angle of the frequency response is obtained as $\angle G(j\omega)$ and represents the phase difference between input and output at each frequency.

## LTI Model for Circadian and Light/Dark Regulation of *CAB2*

The regulation of *CAB2* expression shares some similarity with that of $[Ca^{2+}]_{cyt}$, as it is both circadian-regulated and sensitive to transitions between light and dark. To establish the extent of these similarities from a dynamical systems point of view, we derived an LTI model in the same way as was done for $[Ca^{2+}]_{cyt}$. *CCA1* expression was used as a proxy¶ for circadian regulation subject to a delay $\tau_{CCA1}$, with a binary input variable representing the availability of light. To form the complete input-output estimation dataset, *CAB2::luc* luminescence was measured in *Arabidopsis* seedlings for 2 cycles of 12L/12D before transfer to DD (Fig. S9A), and pre-processing was done exactly as for the $[Ca^{2+}]_{cyt}$ model (Fig. S1A) to implement varying $\tau_{CCA1}$.

A first order model was able to describe the regulation of *CAB2* by the circadian clock and the light signaling network in wild-type plants. Models of complexity up to 3rd

‡Also we must assume $t_2 \neq 0$, as otherwise the system would trivially decouple into two first order equations.

§This is particularly important as any signal can be decomposed into a sum of oscillating functions, namely sines and cosines using a Fourier series decomposition.

¶The interpretation of a proxy is particularly important in this case, as *CAB2* expression is not arrhythmic in *cca1*-1 loss of function mutants (8). We do not claim that the resulting model provides predictive capabilities for mutant strains, though it may sufficiently represent wild-type dynamics in terms of the transfer function from *CCA1* expression to *CAB2* expression.

order were estimated from the 12L/12D–DD input-output data, and performance assessed using a criterion analogous to Eq. 7 with the 12L/12D–DD (Fig. S9A) and 16L/8D (Fig. S9B) datasets. An optimal choice of $\tau_{CCA1}$ (i.e., a value of $\tau_{CCA1}$ which maximized a correlation coefficient $r_w$ defined for the two datasets) was chosen for each combination of $\tau_{\text{light}}$ and $n$ (Fig. S9C). In the majority of cases, a nonzero $\tau_{\text{light}}$ reduced performance. Similarly to the $[Ca^{2+}]_{cyt}$ regulation case, the optimal values of $\tau_{CCA1}$ were all close to 0 h, indicating that the regulation of *CAB2* by the circadian clock does not involve a long pathway. A small increase in $r_w$ performance was observed when increasing $n$ from 1 to 3, though comparing simulations of the models indicated the quantitative increase to be of little consequence (Fig. S9); the optimal-$\tau_{CCA1}$ (0.6 h) 1st order system with $\tau_{\text{light}}$= 0 h was capable of reproducing the dynamics contained within the available data. The resulting candidate 1st order model may be defined in the frequency domain as:

$$\mathcal{L}\{CAB2(t)\} = \begin{pmatrix} G_{CCA1}(s) \\ G_{\text{light}}(s) \end{pmatrix} \mathcal{L}\{\begin{pmatrix} u_{CCA1}(t-0.6) \\ u_{\text{light}}(t) \end{pmatrix}\}, \quad (20)$$

where

$$G_{CCA1}(s) = \frac{2.0510}{s+0.6644}, \quad (21)$$

$$G_{\text{light}}(s) = \frac{1.2669}{s+0.6644} \quad (22)$$

## Linear Systems modeling of Microarray Data

Linear systems were derived for the set of transcripts which exceeded the pMMC–$\beta$ rhythmicity level 0.05 (9). Four microarray datasets were used for model construction, spanning a mixture of growth conditions (Table S2). Of these four datasets, only three represent dynamics in which there are variations in both *CCA1* transcript abundance (input signal representing the circadian clock) and light availability: LDHC, LongLer and ShortLer. First order models were estimated independently from the LDHC, LongLer and ShortLer datasets before selecting the best model of each transcript. A range of input delays were investigated for the clock pathway ($\tau_{CCA1}$= $0, 0.2, \ldots, 8$ h), though the light input pathway was assumed to be fast ($\tau_{\text{light}}$= 0 h). The best model was selected by the weighted correlation coefficient $r_w$ defined above (Eq. 7), subject to comparisons between each of the four microarray datasets in Table S2 and a corresponding simulation. The performance of optimal models derived from each of the three datasets used for estimation, and the best combination of the three, is depicted in Fig. S10A as a cumulative distribution function. Imposing a minimal model performance of $r_w > 0.75$ results in the selection of 1083 models out of the 3503 considered. An illustration of model performance is presented in Fig. S10B, with $r_w$ varying between 0.25 (bottom panels) and 0.97 (top panels). To assess whether *CCA1* is a good gene to use as a proxy for the circadian oscillator, models were also estimated from *TOC1* transcript abundance data using the same method. Indeed, we found that *TOC1*-driven models were outperformed by *CCA1*-driven models in general (Fig. S10B).

Transcripts were classified as clock-dominated, co-regulated or light-dominated, based on the difference in the Bode magnitude of each input pathway. To ensure comparability, *CCA1* transcript abundance data was normalized by dividing all measured values by their corresponding maximum value prior to model estimation, so that both inputs were varying between 0 and 1. As first order models were used to model the transcript abundance data, the Bode phase plot was identical for both input pathways. The considered first order model is described by the equation

$$\frac{dx}{dt}(t) = ax(t) + b_1 u_{CCA1}(t-\tau_{CCA1}) + b_2 u_{\text{light}}(t), \quad (23)$$

where $x$ is the single state variable. Therefore, the transfer function representation is given by

$$X(s) = G_{CCA1}(s)e^{-s\tau_{CCA1}}U_{CCA1}(s) + G_{\text{light}}(s)U_{\text{light}}(s) \quad (24)$$

where $G_{CCA1}(s) = \frac{b_1}{s-a}$ and $G_{\text{light}}(s) = \frac{b_2}{s-a}$.

To classify a transcript as either being clock/light-dominated or co-regulated, we first computed the Bode magnitude as $20\log|G_i(j\omega)|$ for $i = CCA1, \text{light}$. Next, we computed the difference between the Bode magnitude of $G_{CCA1}$ and $G_{\text{light}}$ for each of the 1083 high-performing ($r_w > 0.75$) models (Fig. S10). Since for the model of $[Ca^{2+}]_{cyt}$ there was a difference of 7 dB at $\omega = \frac{\pi}{12}$ (corresponding to a period of 24 h; Fig. 3A) and $[Ca^{2+}]_{cyt}$ is regulated both by the circadian oscillator and light, we chose to use 7 dB as a threshold for classifying all transcripts. We defined a transcript as being clock-dominated if $20\log\frac{|G_{CCA1}(j\omega)|}{|G_{\text{light}}(j\omega)|} > 7\ dB$, light-dominated if $20\log\frac{|G_{CCA1}(j\omega)|}{|G_{\text{light}}(j\omega)|} < -7\ dB$ and co-regulated if $-7\ dB < 20\log\frac{|G_{CCA1}(j\omega)|}{|G_{\text{light}}(j\omega)|} < 7\ dB$.

## Ecotype Variability in the Shape of the $[Ca^{2+}]_{cyt}$ Oscillation in 16L/8D Cycles

AEQ luminescence was measured in a number of *Arabidopsis* accessions under 16L/8D (Fig. S8A). While all accessions had a broad peak in luminescence at mid-day, C24 (7) and WS (10) had an additional sharp peak immediately after dawn. The early peak observed in WS (10) was not present in WS (11). WS ecotypes contain a *phytochrome D* (*phyD*) mutation (12), which may provide an explanation for the variation in dynamics.

1. Ljung, L. (1999) *Systems Identification: Theory for the User*. (Prentice Hall).

2. Locke, J. C. W, Millar, A. J, & Turner, M. S. (2005) Modelling genetic networks with noisy and varied experimental data: the circadian clock in *Arabidopsis thaliana*. *J Theor Biol* **234**, 383–393.

3. Locke, J. C. W, Southern, M. M, Kozma-Bognár, L, Hibberd, V, Brown, P. E, Turner, M. S, & Millar, A. J. (2005) Extension of a genetic network model by iterative experimentation and mathematical analysis. *Mol Syst Biol* **1**, 2005.0013.

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

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

6. Johnson, C. H, Knight, M. R, Kondo, T, Masson, P, Sedbrook, J, Haley, A, & Trewavas, A. (1995) Circadian oscillations of cytosolic and chloroplastic free calcium in plants. *Science* **269**, 1863–1865.

7. Xu, X, Hotta, C. T, Dodd, A. N, Love, J, Sharrock, R, Lee, Y. W, Xie, Q, Johnson, C. H, & Webb, A. A. R. (2007) Distinct light and clock modulation of cytosolic free $Ca^{2+}$ oscillations and rhythmic *CHLOROPHYLL A/B BINDING PROTEIN2* promoter activity in *Arabidopsis*. *Plant Cell* **19**, 3474–3490.

8. Green, R. M & Tobin, E. M. (1999) Loss of the circadian clock-associated protein 1 in *Arabidopsis* results in altered clock-regulated gene expression. *Proc Natl Acad Sci U S A* **96**, 4176–4179.

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

10. Ding, Z, Millar, A. J, Davis, A. M, & Davis, S. J. (2007) *TIME FOR COFFEE* encodes a nuclear regulator in the *Arabidopsis* circadian clock. *Plant Cell* **19**, 1522–1536.

11. Knight, H, Trewavas, A. J, & Knight, M. R. (1997) Calcium signalling in *Arabidopsis thaliana* responding to drought and salinity. *Plant J* **12**, 1067–1078.

12. Aukerman, M. J, Hirschfeld, M, Wester, L, Weaver, M, Clack, T, Amasino, R. M, & Sharrock, R. A. (1997) A deletion in the *PHYD* gene of the arabidopsis wassilewskija ecotype defines a role for phytochrome d in red/far-red light sensing. *Plant Cell* **9**, 1317–1326.

13. Shampine, L. F & Reichelt, M. W. (1997) The MATLAB ODE Suite. *SIAM J. Sci. Comp.* **18**, 1–22.

14. Love, J, Dodd, A. N, & Webb, A. A. R. (2004) Circadian and diurnal calcium oscillations encode photoperiodic information in *Arabidopsis*. *Plant Cell* **16**, 956–966.

15. Nakamichi, N, Kita, M, Ito, S, Sato, E, Yamashino, T, & Mizuno, T. (2005) The *Arabidopsis* pseudo-response regulators, PRR5 and PRR7, coordinately play essential roles for circadian clock function. *Plant Cell Physiol* **46**, 609–619.

16. Kevei, E, Gyula, P, Fehér, B, Tóth, R, Viczián, A, Kircher, S, Rea, D, Dorjgotov, D, Schäfer, E, Millar, A. J, Kozma-Bognár, L, & Nagy, F. (2007) *Arabidopsis thaliana* circadian clock is regulated by the small GTPase LIP1. *Curr Biol* **17**, 1456–1464.

17. Zhou, X & Su, Z. (2007) EasyGO: Gene ontology-based annotation and functional enrichment analysis tool for agronomical species. *BMC Genomics* **8**, 246.

18. Fluhr, R, Kuhlemeier, C, Nagy, F, & Chua, N.-H. (1986) Organ-specific and light-induced expression of plant genes. *Science* **232**, 1106–1112.

19. Lissemore, J. L & Quail, P. H. (1988) Rapid transcriptional regulation by phytochrome of the genes for phytochrome and chlorophyll a/b-binding protein in *Avena sativa*. *Mol Cell Biol* **8**, 4840–4850.

20. Anderson, S. L & Kay, S. A. (1995) Functional dissection of circadian clock- and phytochrome-regulated transcription of the *Arabidopsis CAB2* gene. *Proc Natl Acad Sci U S A* **92**, 1500–1504.

21. Giuliano, G, Hoffman, N. E, Ko, K, Scolnik, P. A, & Cashmore, A. R. (1988) A light-entrained circadian clock controls transcription of several plant genes. *EMBO J* **7**, 3635–3642.

22. Taylor, W. C. (1989) Transcriptional regulation by a circadian rhythm. *Plant Cell* **1**, 259–264.

23. Millar, A. J & Kay, S. A. (1991) Circadian control of *cab* gene transcription and mRNA accumulation in *Arabidopsis*. *Plant Cell* **3**, 541–550.

24. Zhong, H. H, Young, J. C, Pease, E. A, Hangarter, R. P, & McClung, C. R. (1994) Interactions between light and the circadian clock in the regulation of *CAT2* expression in *Arabidopsis*. *Plant Physiol* **104**, 889–898.

25. Anderson, J. W & Done, J. (1978) Light-dependent assimilation of nitrite by isolated pea chloroplasts. *Plant Physiol* **61**, 692–697.

26. Pilgrim, M. L, Caspar, T, Quail, P. H, & McClung, C. R. (1993) Circadian and light-regulated expression of nitrate reductase in *Arabidopsis*. *Plant Mol Biol* **23**, 349–364.

27. Tucker, D. E, Allen, D. J, & Ort, D. R. (2004) Control of nitrate reductase by circadian and diurnal rhythms in tomato. *Planta* **219**, 277–285.

28. Sen, S. P & Leopold, A. C. (1956) Influence of light and darkness upon carbon dioxide fixation. *Plant Physiol* **31**, 323–329.

29. Tolbert, N. E & Gailey, F. B. (1955) Carbon dioxide fixation by etiolated plants after exposure to white light. *Plant Physiol* **30**, 491–499.

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

31. Hennessey, T. L & Field, C. B. (1991) Circadian rhythms in photosynthesis : Oscillations in carbon assimilation and stomatal conductance under constant conditions. *Plant Physiol* **96**, 831–836.

32. Koller, D. (1990) Light-driven leaf movements. *Plant, Cell and Environment* **13**, 615–632.

33. Nishizaki, Y. (1996) Effects of blue light on electrical potential and turgor in pulvinar motor cells of *Phaseolus*. *Journal of Plant Research* **109**, 93–97.

34. Alford, D. K & Tibbitts, T. W. (1970) Circadian rhythm of leaves of *Phaseolus angularis* plants grown in a controlled carbon dioxide and humidity environment. *Plant Physiol* **46**, 99–102.

35. Rubinstein, B. (1971) Auxin and red light in the control of hypocotyl hook opening in beans. *Plant Physiol* **48**, 187–192.

36. Samimy, C. (1978) Effect of light on ethylene production and hypocotyl growth of soybean seedlings. *Plant Physiol* **61**, 772–774.

37. Dowson-Day, M. J & Millar, A. J. (1999) Circadian dysfunction causes aberrant hypocotyl elongation patterns in *Arabidopsis*. *Plant J* **17**, 63–71.

38. Nozue, K, Covington, M. F, Duek, P. D, Lorrain, S, Fankhauser, C, Harmer, S. L, & Maloof, J. N. (2007) Rhythmic growth explained by coincidence between internal and external cues. *Nature* **448**, 358–361.

39. Robson, P. R. H, Whitelam, G. C, & Smith, H. (1993) Selected components of the shade-avoidance syndrome are displayed in a normal manner in mutants of *Arabidopsis thaliana* and *Brassica rapa* deficient in phytochrome B. *Plant Physiol* **102**, 1179–1184.

40. Salter, M. G, Franklin, K. A, & Whitelam, G. C. (2003) Gating of the rapid shade-avoidance response by the circadian clock in plants. *Nature* **426**, 680–683.

41. Ku, S.-B, Edwards, G. E, & Tanner, C. B. (1977) Effects of light, carbon dioxide, and temperature on photosynthesis, oxygen inhibition of photosynthesis, and transpiration in *Solanum tuberosum*. *Plant Physiol* **59**, 868–872.

42. Assmann, S. M. (1988) Enhancement of the stomatal response to blue light by red light, reduced intercellular concentrations of $CO_2$, and low vapor pressure differences. *Plant Physiol* **87**, 226–231.

43. Somers, D. E, Devlin, P. F, & Kay, S. A. (1998) Phytochromes and cryptochromes in the entrainment of the *Arabidopsis* circadian clock. *Science* **282**, 1488–1490.

44. Gorton, H. L, Williams, W. E, Binns, M. E, Gemmell, C. N, Leheny, E. A, & Shepherd, A. C. (1989) Circadian stomatal rhythms in epidermal peels from *Vicia faba*. *Plant Physiol* **90**, 1329–1334.

45. Reid, H. B, Moore, P. H, & Hamner, K. C. (1967) Control of flowering of *Xanthium pensylvanicum* by red and far-red light. *Plant Physiol* **42**, 532–540.

46. Deitzer, G. F, Hayes, R. G, & Jabben, M. (1982) Phase shift in the circadian rhythm of floral promotion by far red energy in *Hordeum vulgare* l. *Plant Physiol* **69**, 597–601.

47. Devlin, P. F, Halliday, K. J, Harberd, N. P, & Whitelam, G. C. (1996) The rosette habit of *Arabidopsis thaliana* is dependent upon phytochrome action: novel phytochromes control internode elongation and flowering time. *Plant J* **10**, 1127–1134.

48. Halaban, R. (1968) The flowering response of coleus in relation to photoperiod and the circadian rhythm of leaf movement. *Plant Physiol* **43**, 1894–1898.

49. Hicks, K. A, Millar, A. J, Carré, I. A, Somers, D. E, Straume, M, Meeks-Wagner, D. R, & Kay, S. A. (1996) Conditional circadian dysfunction of the *Arabidopsis early-flowering 3* mutant. *Science* **274**, 790–792.

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

51. Baum, G, Long, J. C, Jenkins, G. I, & Trewavas, A. J. (1999) Stimulation of the blue light phototropic receptor NPH1 causes a transient increase in cytosolic $Ca^{2+}$. *Proc Natl Acad Sci U S A* **96**, 13554–13559.

52. Neuhaus, G, Bowler, C, Hiratsuka, K, Yamagata, H, & Chua, N. H. (1997) Phytochrome-regulated repression of gene expression requires calcium and cGMP. *EMBO J* **16**, 2554–2564.

53. Wu, Y, Hiratsuka, K, Neuhaus, G, & Chua, N. H. (1996) Calcium and cGMP target distinct phytochrome-responsive elements. *Plant J* **10**, 1149–1154.

54. Michael, T. P, Mockler, T. C, Breton, G, McEntee, C, Byer, A, Trout, J. D, Hazen, S. P, Shen, R, Priest, H. D, Sullivan, C. M, Givan, S. A, Yanovsky, M, Hong, F, Kay, S. A, & Chory, J. (2008) Network discovery pipeline elucidates conserved time-of-day-specific cis-regulatory modules. *PLoS Genet* **4**, e14.

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

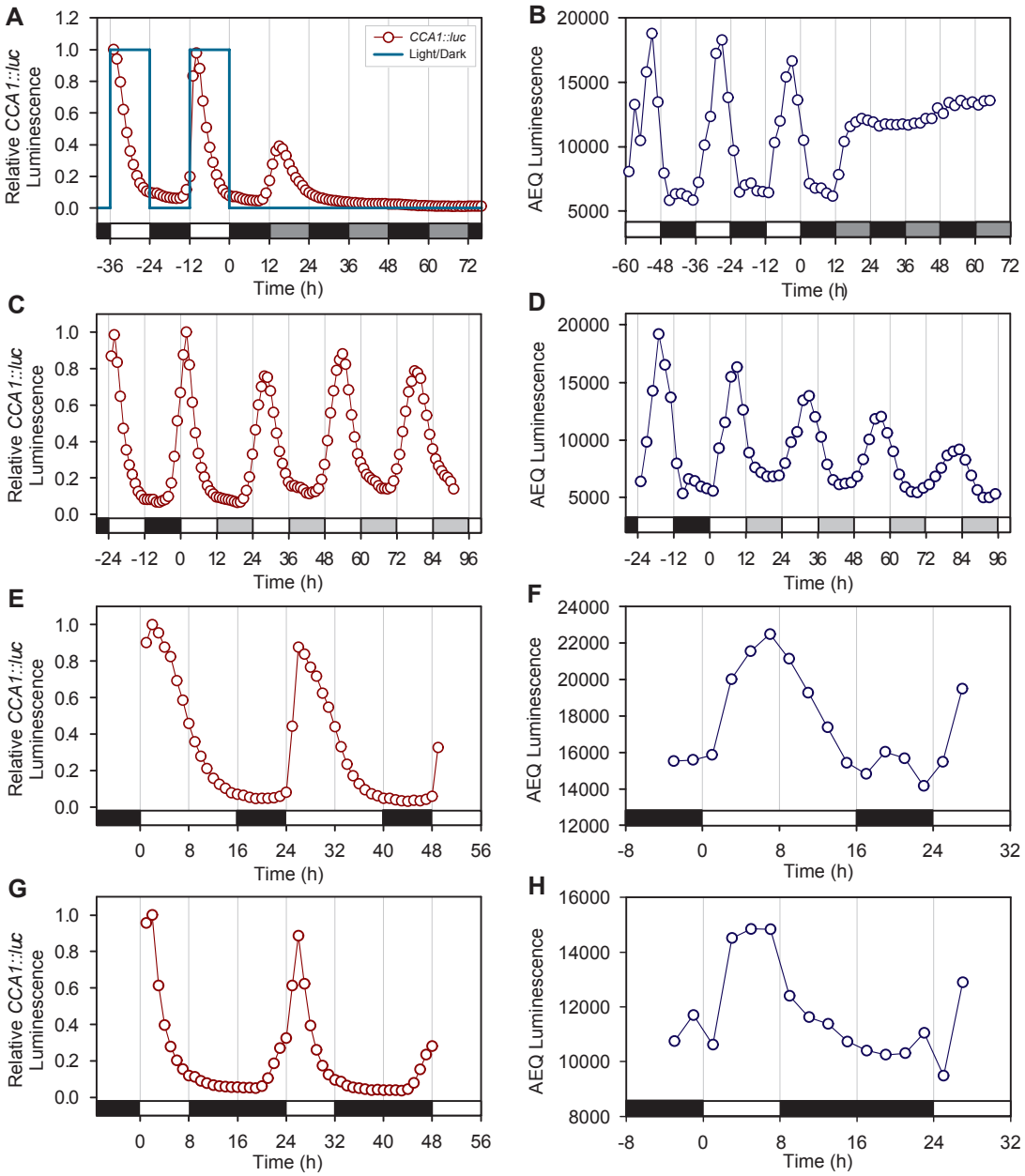


**Figure S1: Pre-processing the input-output data prior to model estimation.**
(A,C,E,G) Relative *CCA1::luc* luminescence quantified for 10 day old *Arabidopsis* seedlings, where open red circles represent mean average of photon counts per seedling cluster integrated for 800 s. In (A), the binary variable representing the availability of light (light=1, dark=0) corresponding to the experimental photoperiodic conditions is represented by a solid blue line. (B,D,F,H) AEQ luminescence quantified for 10 day old *Arabidopsis* seedlings, where open blue circles represent mean average of photon counts per seedling cluster integrated for 1500 s. Bars on abscissa represent light availability; white indicates light, black indicates dark, dark gray is subjective day in DD and light gray is subjective night in LL.

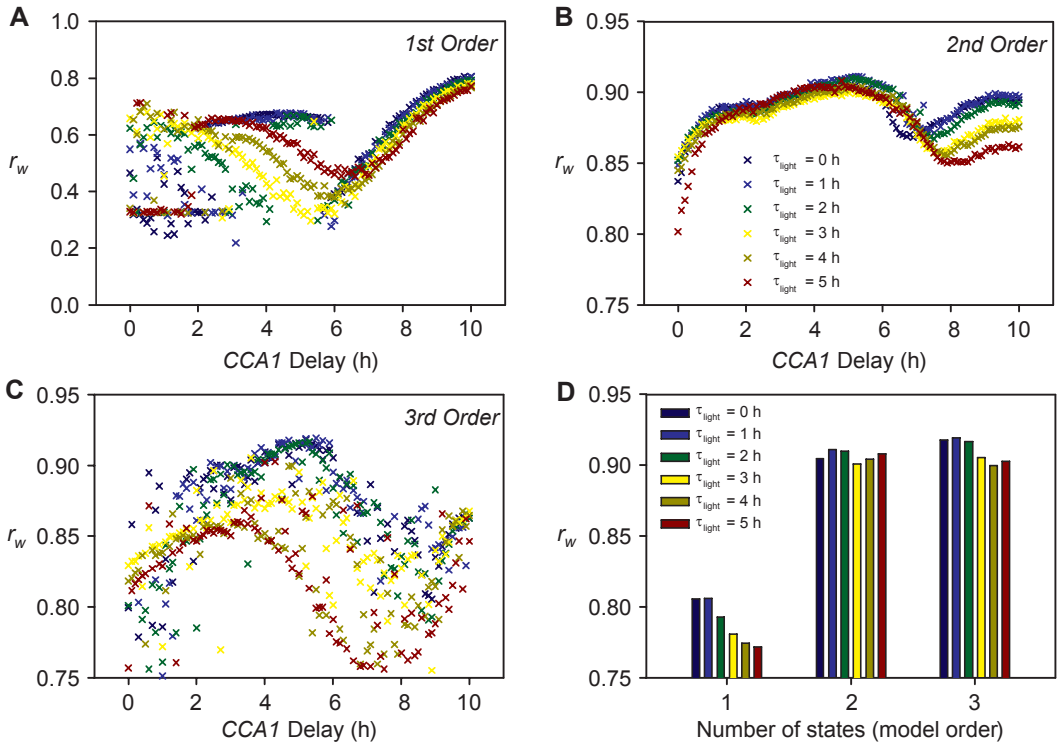


**Figure S2: Evaluation of the weighted cross-validation function for models of circadian regulation of $[Ca^{2+}]_{cyt}$.**
Performance function $r_w$ expressed as a function of $\tau_{CCA1}$, for (A) 1st, (B) 2nd, and (C) 3rd order models. Values of $\tau_{light}$ are indicated by different colors (see legend of (D)). (D) Summary of model performance for each combination of $\tau_{light}$ and $n$, where bar heights correspond to $r_w$ values for the best choice of $\tau_{CCA1}$.

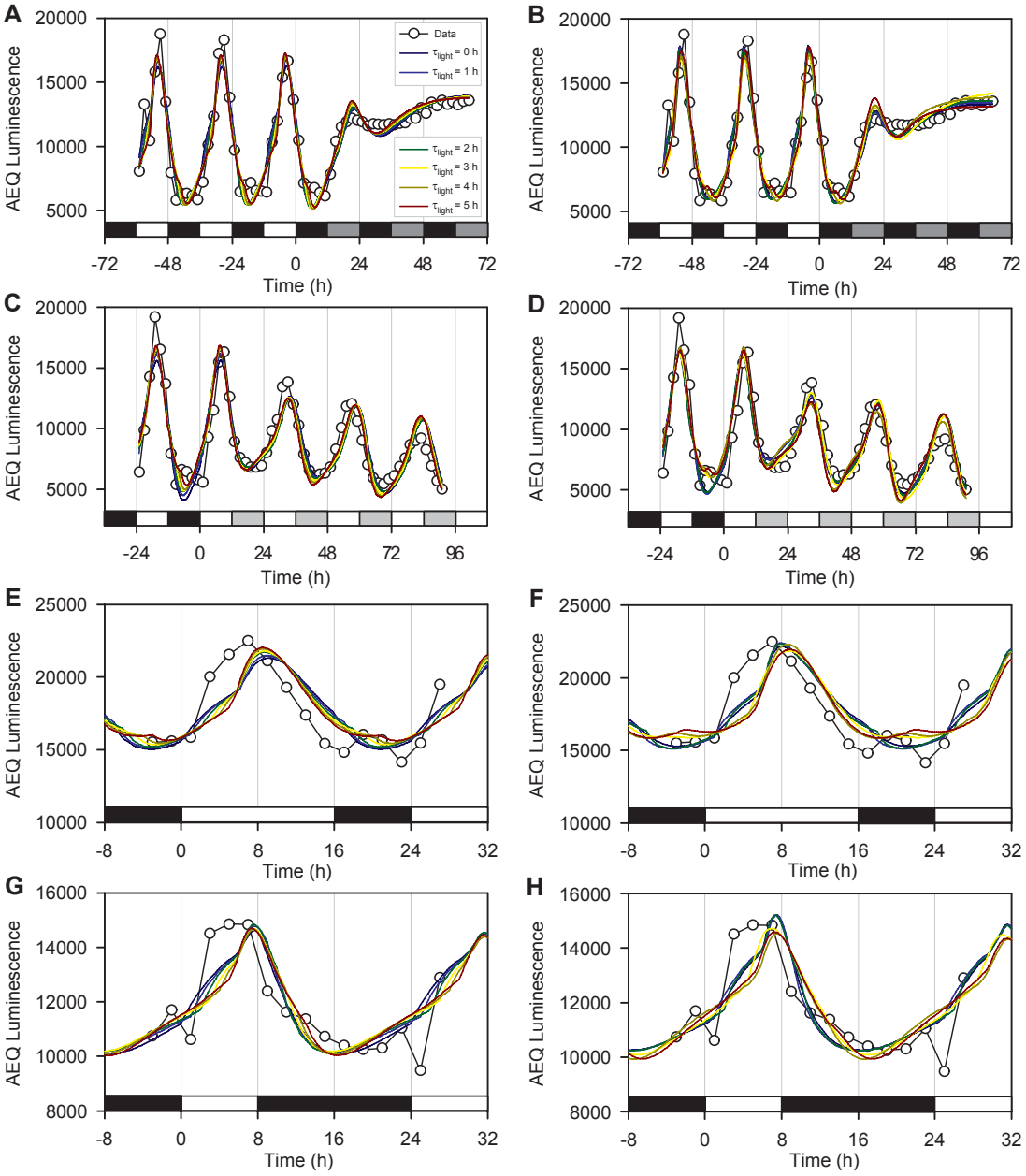


**Figure S3: Cross-validation of (A,C,E,G) 2nd and (B,D,F,H) 3rd order models of circadian regulation of $[Ca^{2+}]_{cyt}$.**
Linear state-space model equations were solved numerically with MATLAB™'s *ode23* solver (solid lines) (13). Simulation results were compared with corresponding experimental data for (A,B) 12L/12D cycles before transfer to DD, (C,D) 12L/12D cycles before transfer to LL, (E,F) 16L/8D cycles, and (G,H) 8L/16D cycles. For each $\tau_{\text{light}} = 0, 1, \ldots, 5$, the models were simulated for the value of $\tau_{CCA1}$ which had the highest value of $r_w$, the weighted sum of sample correlation coefficients (see Fig. S2). Simulated model outputs have been scaled to overlap with AEQ luminescence data (open circles) by finding the best linear scaling of the form $a\hat{y} + b$, where $\hat{y}$ is the simulated output. The *polyfit* function in MATLAB™ was used to evaluate $a$ and $b$ as minimizers of $\sum_i^{N_d} (y^{(i)} - a\hat{y}^{(i)} - b)^2$ where $y^{(i)}$ are the measured AEQ luminescence values, $\hat{y}^{(i)}$ are the corresponding simulated model output values and $N_d$ is the number of data values. Bars on abscissa represent light regime in each experiment/simulation. White indicates light, black indicates dark, light gray is light in the subjective night and dark gray is dark in subjective daytime.

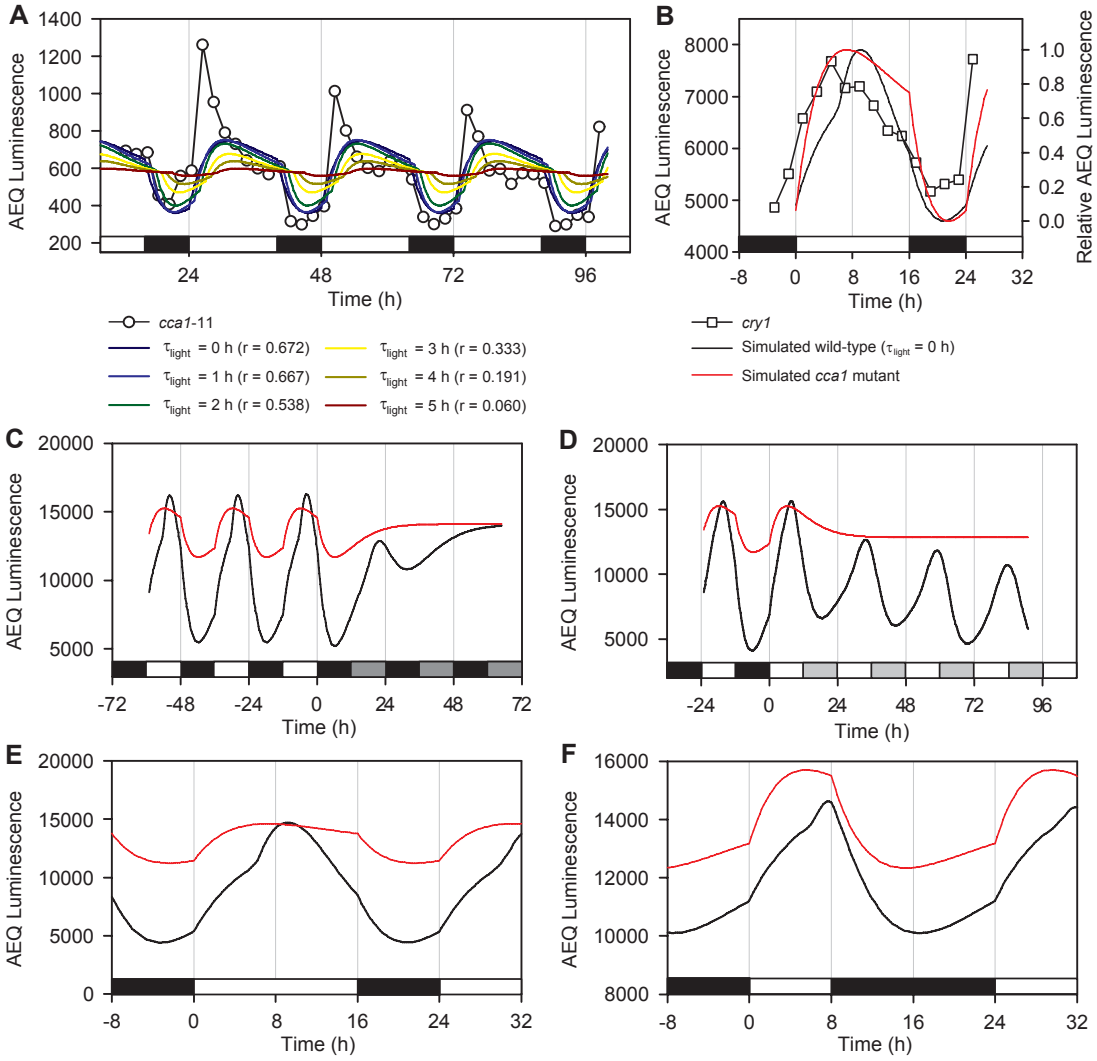


**Figure S4: Simulation of the response of $[Ca^{2+}]_{cyt}$ to a mutation in *CCA1*.**
(A) AEQ luminescence measured by photon counting imaging in 10 day old *cca1*-11 mutant *Arabidopsis* seedlings entrained to 16L/8D cycles (open circles). Superimposed on the same plot are equivalent simulation results for AEQ luminescence obtained from optimal-$\tau_{CCA1}$ 2$^{nd}$ order models for different choices of $\tau_{light}$ (solid lines). These simulation results have been obtained after normalization, which was performed to overlap with the measured data in the least-squares sense. The corresponding correlation coefficient values are indicated in the associated legend. (B) AEQ luminescence measured in 10 day old *cry1* mutant *Arabidopsis* seedlings entrained to 16L/8D cycles (open squares), compared with simulation results for relative AEQ luminescence in wild-type (black line) and *cca1* mutants (red line); these results were obtained using the optimal-$\tau_{CCA1}$ model for $\tau_{light}$= 0 h. Wild-type and *cca1* mutant simulations are compared for (C) 12L/12D cycles before DD, (D) 12L/12D cycles before LL, (E) 16L/8D cycles, and (F) 8L/16D cycles, without normalization. In all panels, *cca1* mutants are simulated by setting $u_{CCA1}$ equal to zero throughout. Linear state-space model equations were solved numerically with MATLAB$^{TM}$'s *ode23* solver (13). For each $\tau_{light} = 0, 1, \ldots, 5$, the models were simulated for the optimal choice of $\tau_{CCA1}$, defined as the minimizer for the weighted sum of correlations with cross-validation test data, $r_w$ (Eq. 6). Bars on abscissa represent light regime in each simulation. White indicates light, black indicates dark, light gray in (D) is light in the subjective night and dark gray in (C) is dark in subjective daytime.

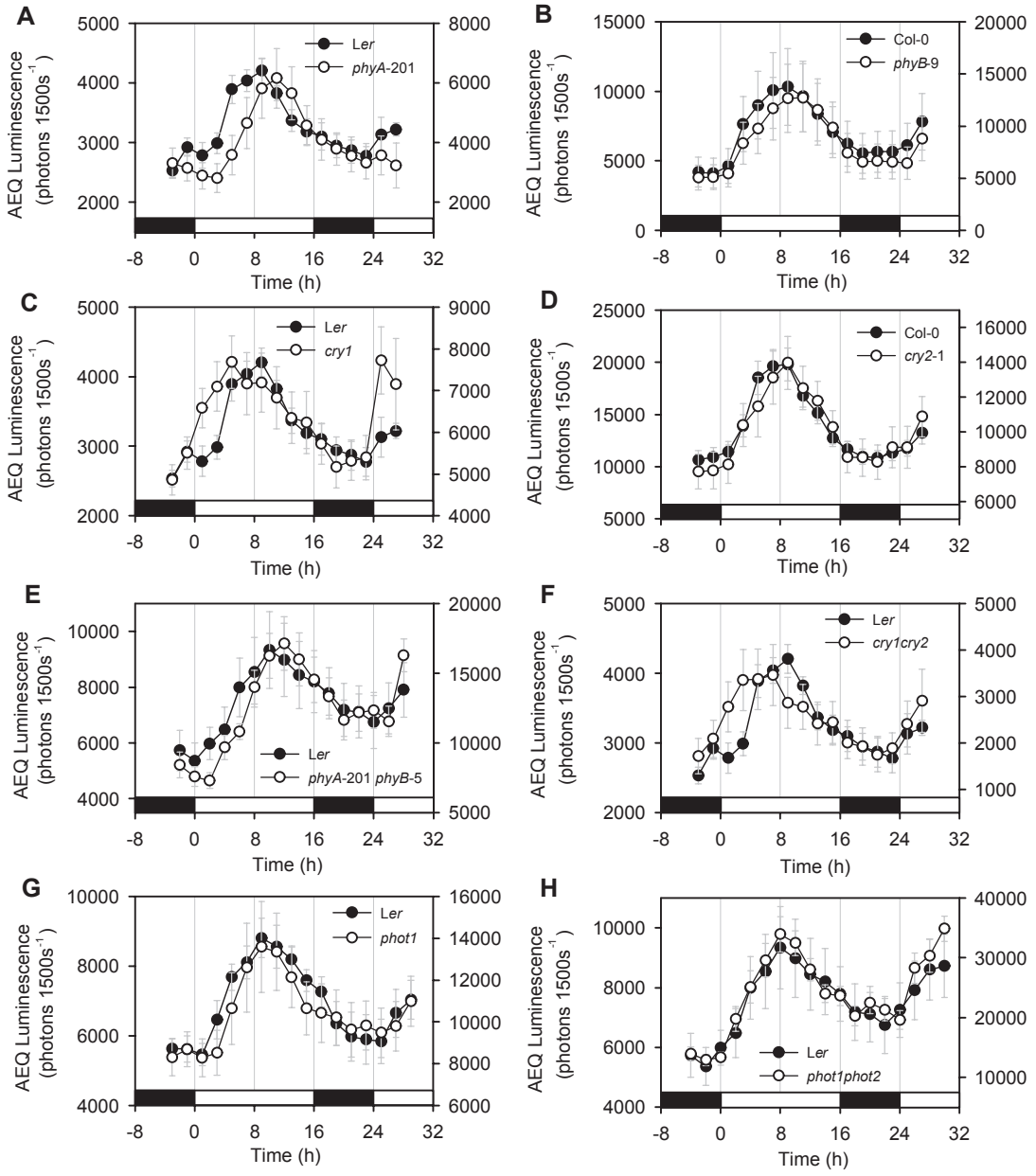


**Figure S5: Effects of photoreceptor mutations on the $[Ca^{2+}]_{cyt}$ oscillation in 16L/8D cycles.**
In all plots, AEQ luminescence is quantified for 10 day old seedlings entrained and imaged in 16L/8D cycles (100 $\mu$mol $m^{-2}$ $s^{-1}$). The corresponding traces representative of the $[Ca^{2+}]_{cyt}$ oscillation in (A) *phyA*-201, (B) *phyB*-9, (C) *cry1*, (D) *cry2*-1, (E) *phyA*-201*phyB*-5, (F) *cry1cry2*, (G) *phot1*, and (H) *phot1phot2* are shown; wildtype seedlings are represented by closed circles; the mutant line is indicated by open circles. In all figures, open bars represent light, closed bars darkness. Luminescence values on the left hand side of the figure indicate those of the wildtype, values on the right those of the mutant. Values are means, bars represent the standard error of the mean, $n = 3 - 5$.

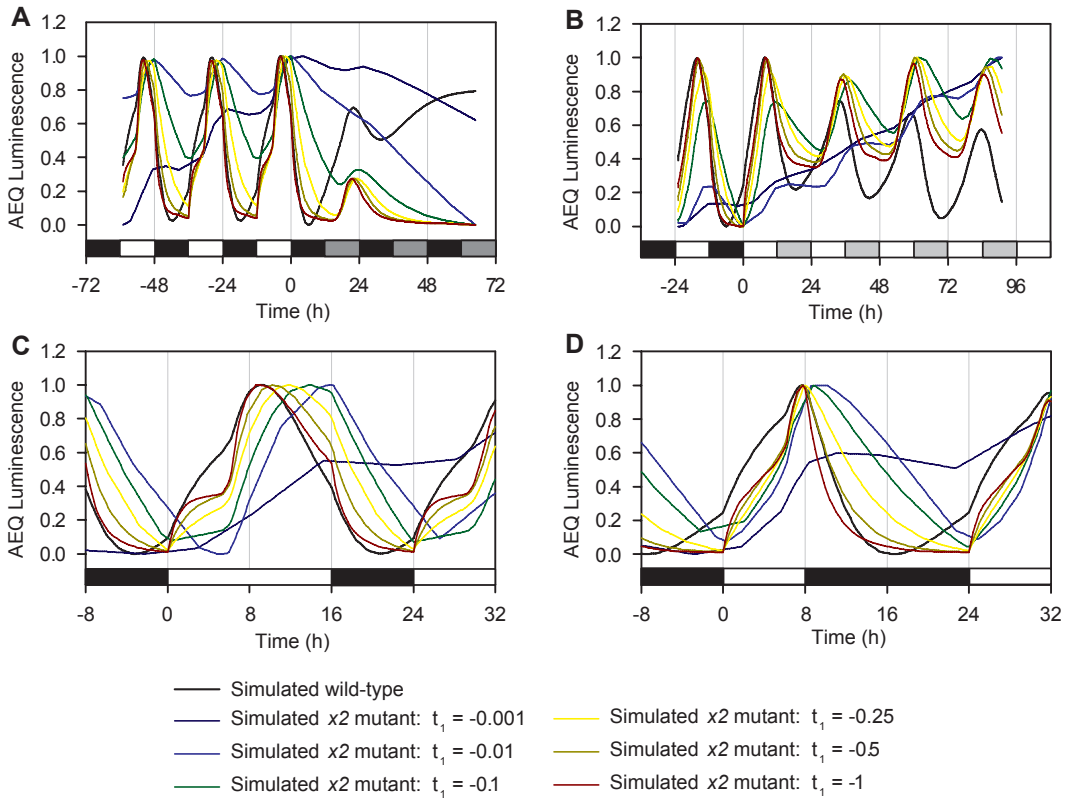


**Figure S6: Simulation of a mutation in the $[Ca^{2+}]_{cyt}$model hidden variable *X2*.**
Simulations of (A) LD–DD, (B) LD–LL, (C) 16L/8D and (D) 8L/16D cycles reflect removal of the hidden variable *X2*, by solving the first order equation $\dot{x}(t) = t_1 x(t) + b_{11} u_{CCA1}(t - \tau_{CCA1}) + b_{12} u_{\text{light}}(t)$. $t_1$ is the free parameter from a similarity transformation which defines the family of mutant responses. $b_{11}$ and $b_{12}$ are the entries in the first row of the *B* matrix, estimated from wild-type data using the *pem* prediction-error minimization function in MATLAB$^{TM}$ (Systems Identification Toolbox) (1). The estimated model corresponds to the case where $\tau_{\text{light}} = 0$ h, and $\tau_{CCA1} = 5.2$ h. The *CCA1*-dependent pathway is assumed to be unaffected by the mutation, as the wild-type data (Fig. S1 B,D,F,H) is used for these simulations. The first 24 hours of input data is cycled through 5 times to remove transitory behaviors. Linear state-space model equations were solved numerically with MATLAB$^{TM}$'s *ode23* solver (13). Bars on abscissa represent light regime in each simulation. White indicates light, black indicates dark, light gray (B) is light in the subjective night and dark gray (A) is dark in subjective daytime.

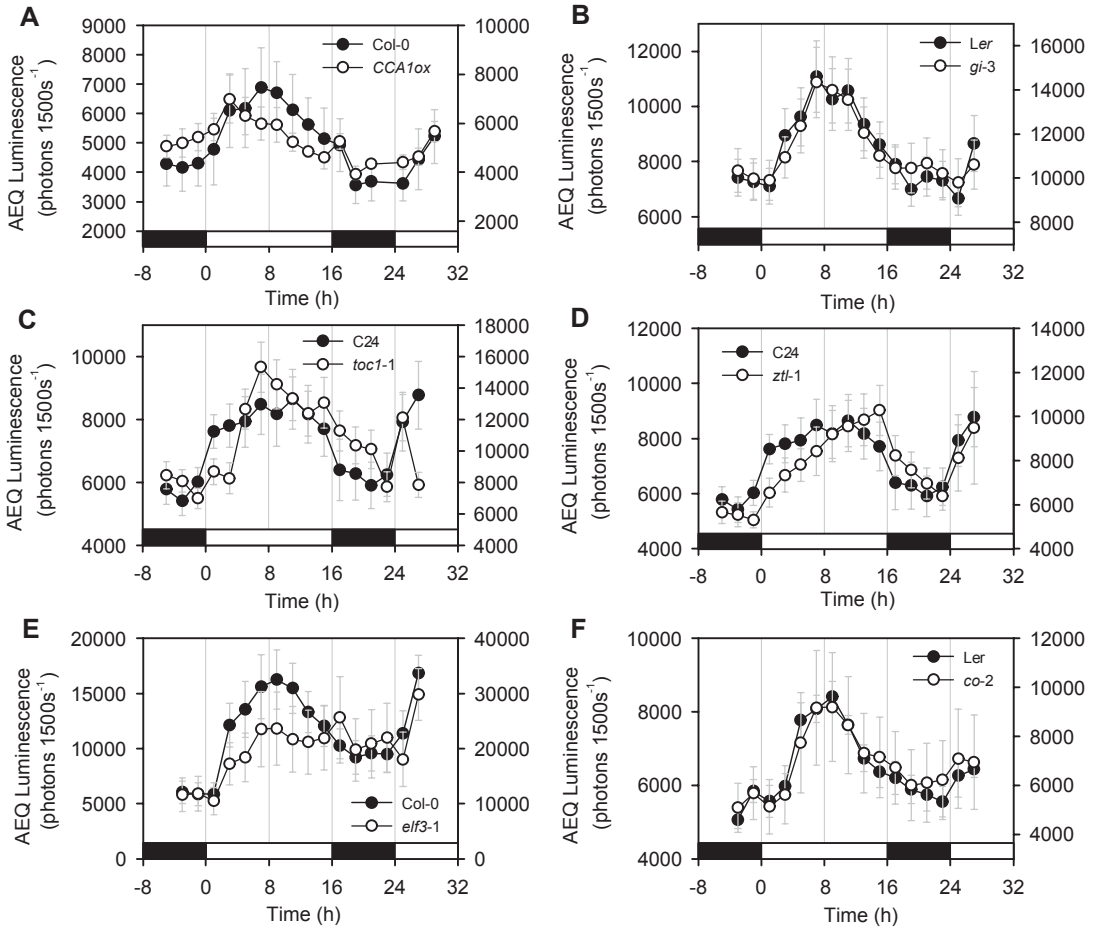


**Figure S7: Effects of mutations associated with the central oscillator and photoperiodic flowering time pathways on the $[Ca^{2+}]_{cyt}$ oscillation in 16L/8D cycles.**

In these plots, AEQ luminescence is quantified for 10 day old seedlings entrained and imaged in 16L/8D cycles (100 $\mu$mol m$^{-2}$ s$^{-1}$). The corresponding traces representative of the $[Ca^{2+}]_{cyt}$ oscillation in (A) *CCA1ox*, (B) *gi*-3 , (C) *toc1*-1, (D) *ztl*-1, (E) *elf3*-1 and (F) *co*-2 are shown; wildtype seedlings are represented by closed circles; the mutant line is indicated by open circles. In all figures, open bars represent light, closed bars darkness. Luminescence values on the left hand side of the figure indicate those of the wildtype, values on the right those of the mutant. Values are means, bars represent the standard error of the mean, $n = 3 - 5$.

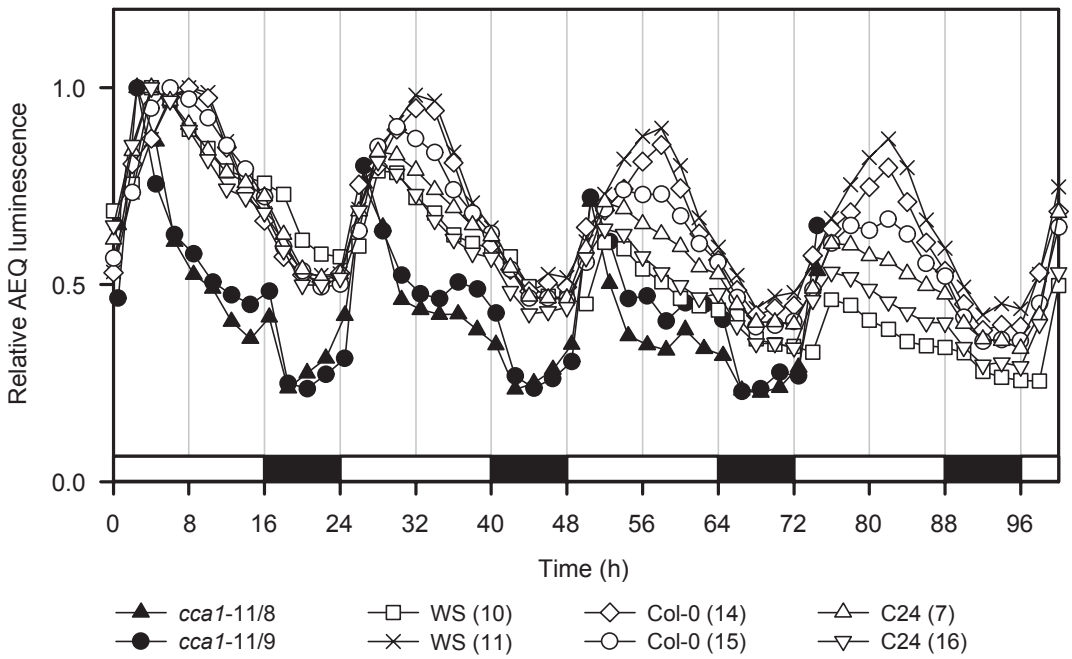


**Figure S8: Ecotype variability in the shape of the $[Ca^{2+}]_{cyt}$ oscillation in 16L/8D cycles.**
AEQ luminescence quantified for 10 day old seedlings entrained and imaged in 16L/8D cycles (100 $\mu$mol $m^{-2}$ $s^{-1}$) for wild-type Col-0, C24 and WS accessions, and *cca1*-11 null mutants. WS (10) is the wild-type control which corresponds to *cca1*-11 mutants. The data in the graph are averages of two independent experiments. Data for two independent transformants of the *cca1*-11 null mutant expressing AEQ, *cca1*-11/8 and *cca1*-11/9, is shown. In abscissa, open bars represent light, closed bars darkness. In the key and this figure legend, bracketed numbers are bibliographical references which indicate the source of the measured ecotype.

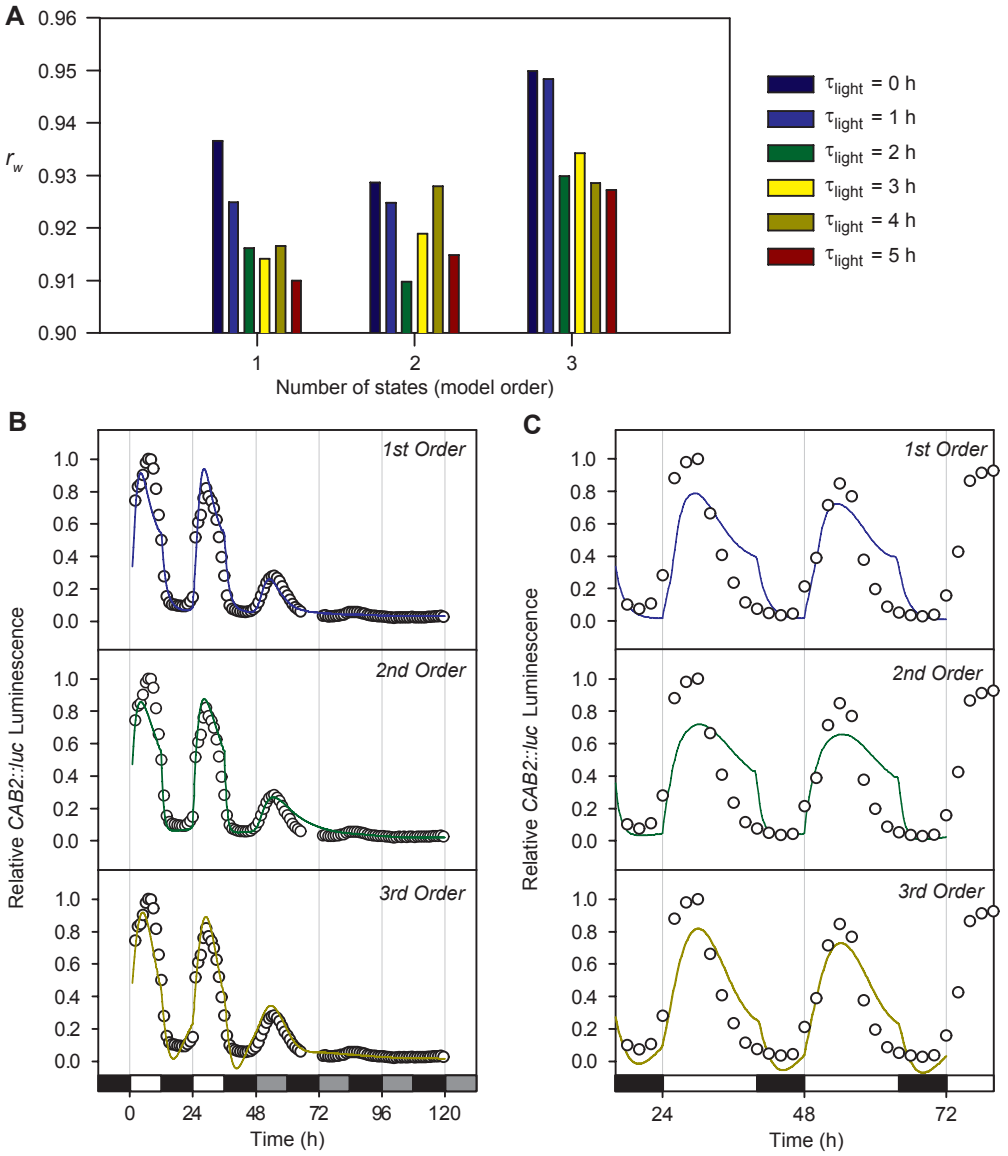


**Figure S9: Identification and selection of *CAB2* models.**
(A) Model performance for optimal-$\tau_{CCA1}$ as a function of model order (i.e., $n$=number of states) and $\tau_{\text{light}}$. (B,C) Linear state-space model equations solved numerically with MATLAB$^{\text{TM}}$'s *ode23* solver (solid lines) in (A) LD–LL and (B) 16L/8D cycles (13). For $\tau_{\text{light}} = 0$, the models were simulated for the value of $\tau_{CCA1}$ which had the highest value of $r_w$. Comparisons are shown for models of increasing complexity from the top panel (1$^{\text{st}}$ order) to the bottom panel (3$^{\text{rd}}$ order). Simulated model outputs were scaled to overlap with *CAB2::luc* luminescence data (open circles) by finding the best linear scaling of the form $a\hat{y} + b$, where $\hat{y}$ is the simulated output. The *polyfit* function in MATLAB$^{\text{TM}}$ was used to evaluate $a$ and $b$ as minimizers of $\sum_i^{N_d}(y^{(i)} - a\hat{y}^{(i)} - b)^2$ where $y^{(i)}$ are the measured *CAB2::luc* luminescence values, $\hat{y}^{(i)}$ are the corresponding simulated model output values and $N_d$ is the number of data values. Bars on abscissa represent 16L/8D cycle light regime, white indicating light, black indicating dark and light gray indicating dark during subjective day.

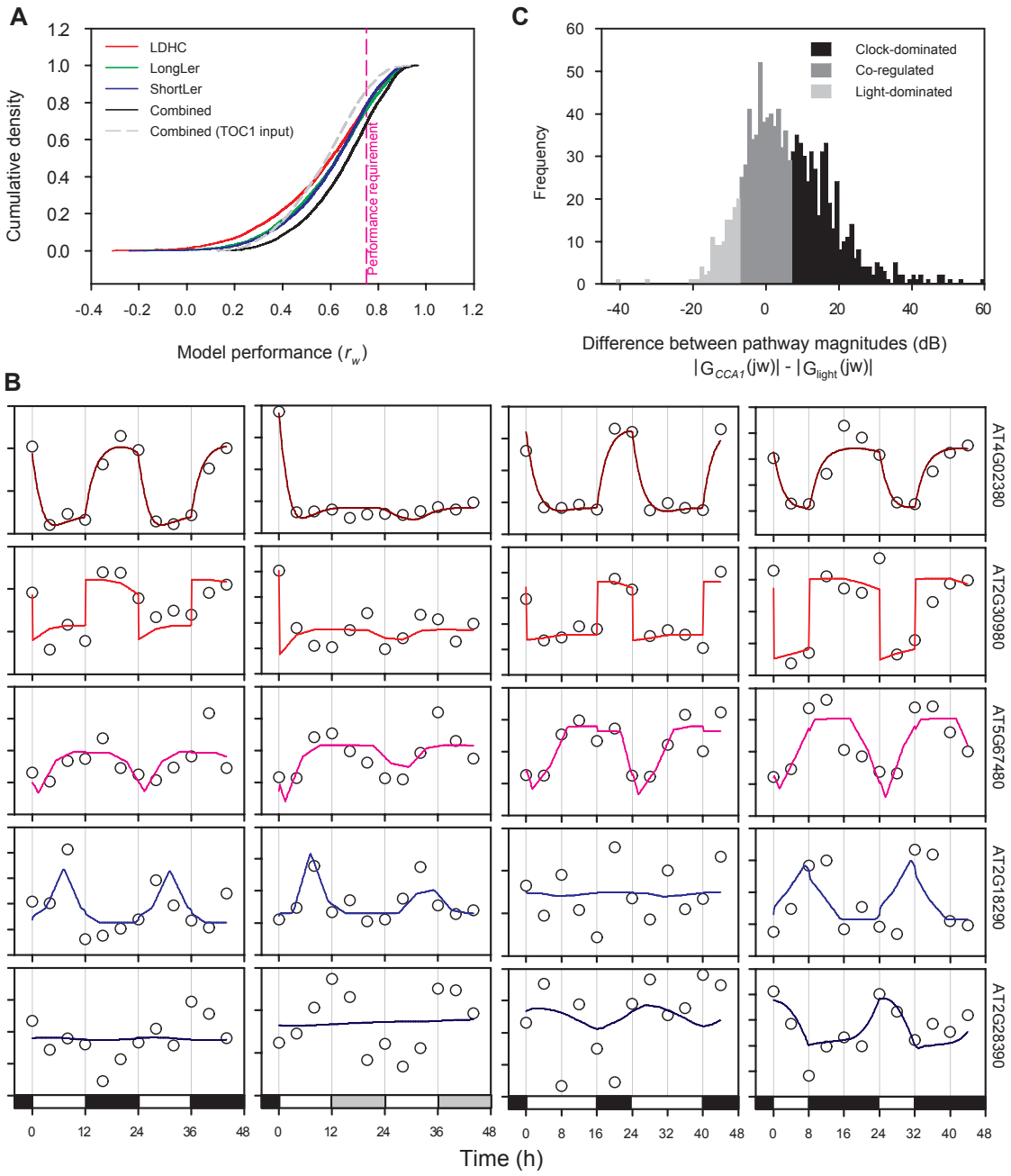


**Figure S10: Mathematical modeling of rhythmic transcripts from microarray data.**
(A) First order delay linear systems estimated from three different datasets (LDHC, LongLer and ShortLer) and analyzed for performance using $r_w$. Selection of the best model for each transcript produces the combined model set. Each model set is expressed as a cumulative density function of the model performance. A performance threshold of 0.75 was chosen as the cut-off value for further analysis. (B) Simulation of a range of first order models with highest observed performance (top), followed by $r_w = 0.9, 0.75, 0.5$ and 0.25 (bottom). The four tests which comprize the performance metric are arranged in columns: 12L/12D (left), 12L/12D–LL, 16L/8D and 8L/16D (right). (C) Histograph of difference between the pathway magnitudes of high performance models ($r_w > 0.75$). The models are categorized as clock- or light-dominated if the difference exceeds 7 dB, and co-regulated otherwize.

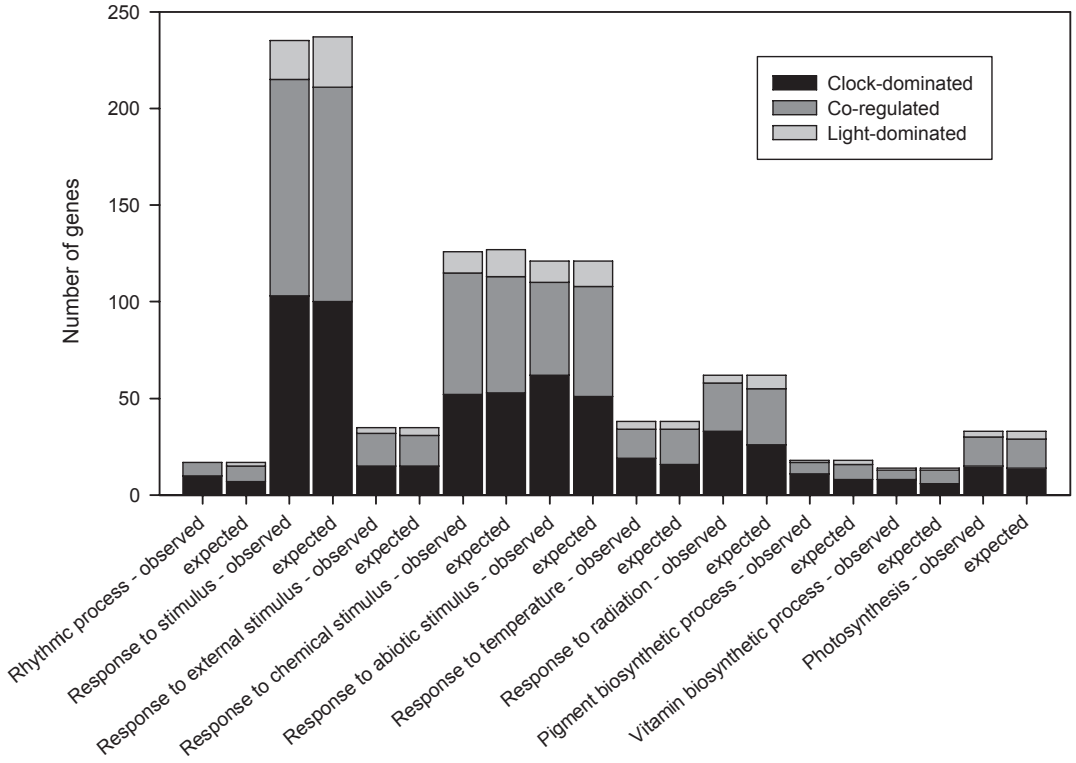


**Figure S11: Gene Ontology analysis of transcript regulation classes.**
The involvement of clock-dominated, co-regulated and light-dominated transcripts in biological processes that are over-represented within the subset of circadian-regulated transcripts (with performance threshold $r_w > 0.75$) was analysed. For each gene ontology (GO) term, the number of genes observed in each of the 3 gene lists is shown and compared to the number of genes expected in each list by chance. The web-based enrichment analysis tool EasyGo was used for the GO term analysis (http://bioinformatics.cau.edu.cn/easygo/) (17).

Table S1: Physiological processes in plants regulated by both the circadian clock and light.

| Process | Regulatory System | |
|---|---|---|
| | Light | Circadian |
| Expression of chlorophyll a/b binding protein (CAB) | 18–20 | 21–23 |
| Expression of catalase 2 (CAT2) | 24 | 24 |
| Nitrate assimilation | 25, 26 | 26, 27 |
| Carbon fixation | 28, 29 | 30, 31 |
| Leaf angle (legumes) | 32, 33 | 34 |
| Hypocotyl elongation | 35, 36 | 37, 38 |
| Shade avoidance | 39 | 40 |
| Stomatal aperture | 41, 42 | 30, 43, 44 |
| Photoperiodic control of flowering time | 45–47 | 48–50 |
| $[Ca^{2+}]_{cyt}$ | 7, 51–53 | 6, 7, 14 |

Table S2: Microarray experiments used for model construction.

| Name | Condition | Age (d) | Accession | Media | Reference |
|---|---|---|---|---|---|
| LDHC | 12 h light at 22°C and 12 h dark at 12°C | 7 | Col-0 | agar, no suc | 54 |
| LongLer | 16 h light, 8 h dark at 22° | 7 | L*er* | agar, 3% suc | 55 |
| ShortLer | 8 h light, 16 h dark at 22° | 7 | L*er* | agar, 3% suc | 55 |
| LL-LDHC | Constant light at 22° (after transfer from LDHC) | 9 | Col-0 | agar, no suc | 54 |

Table S3: Distribution of the transcripts according to dominant regulatory pathway. Transcripts are classified as co-regulated if the difference in magnitude is less than 7 dB. The corresponding numbers of transcripts where oscillation phase information is available for entrained 8L/16D and 16L/8D cycles are given in brackets.

| | Clock-dominated | Co-regulated | Light-dominated | Total |
|---|---|---|---|---|
| All transcripts | 1028 (719) | 1937 (914) | 358 (212) | 3503 (1845) |
| $r_w > 0.75$ | 455 (413) | 507 (381) | 121 (98) | 1083 (892) |

Table S4: Relationship between the dominant regulatory pathway and the rephasing of maximal expression between 8L/16D and 16L/8D. Data are expressed as the % of transcripts within each class that re-phase between 8L/16D and 16L/8D. Asterisks (**) indicate statistical significance at the 95% confidence

| | Clock-dominated | Co-regulated | Light-dominated | All structures | $\chi^2$ statistic |
|---|---|---|---|---|---|
| All transcripts | 24.9% | 55.6% | 54.2% | 43.4% | $9.0 \times 10^{-4}$ ** |
| $r_w > 0.75$ | 17.6% | 47.9% | 54.6% | 35.2% | $6.2 \times 10^{-6}$ ** |