Defining the robust behaviour of the plant clock gene circuit with absolute RNA timeseries and open infrastructure


Digital Object Identifier (DOI): 10.1098/rsob.150042

Link: Link to publication record in Edinburgh Research Explorer

Document Version: Publisher's PDF, also known as Version of record

Published In: Open Biology

Publisher Rights Statement: C 2015 The Authors. Published by the Royal Society under the terms of the Creative Commons Attribution License http://creativecommons.org/licenses/by/4.0/, which permits unrestricted use, provided the original author and source are credited.

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Defining the robust behaviour of the plant clock gene circuit with absolute RNA timeseries and open infrastructure

Anna Flis¹, Aurora Piñas Fernández²†, Tomasz Zielinski², Virginie Mengin¹, Ronan Sulpice¹‡, Kevin Stratford³, Alastair Hume²–³, Alexandra Pokhilko²–⁴, Megan M. Southern⁵, Daniel D. Seaton², Harriet G. McWatters², Mark Stitt¹, Karen J. Halliday² and Andrew J. Millar²

¹Max Planck Institute of Molecular Plant Physiology, Am Muehlenberg 1, 14476 Potsdam-Golm, Germany
²SynthSys and School of Biological Sciences, University of Edinburgh, C.H. Waddington Building, Edinburgh EH9 3JZ, UK
³EPCC, University of Edinburgh, James Clerk Maxwell Building, Edinburgh EH9 3JZ, UK
⁴Institute of Molecular Cell and Systems Biology, University of Glasgow, Bower Building, Glasgow G12 8QQ, UK
⁵Department of Biological Sciences, University of Warwick, Coventry CV4 7AL, UK

Defining the robust behaviour of the plant clock gene circuit with absolute RNA timeseries and open infrastructure

Anna Flis¹, Aurora Piñas Fernández²†, Tomasz Zielinski², Virginie Mengin¹, Ronan Sulpice¹‡, Kevin Stratford³, Alastair Hume²–³, Alexandra Pokhilko²–⁴, Megan M. Southern⁵, Daniel D. Seaton², Harriet G. McWatters², Mark Stitt¹, Karen J. Halliday² and Andrew J. Millar²

¹Max Planck Institute of Molecular Plant Physiology, Am Muehlenberg 1, 14476 Potsdam-Golm, Germany
²SynthSys and School of Biological Sciences, University of Edinburgh, C.H. Waddington Building, Edinburgh EH9 3JZ, UK
³EPCC, University of Edinburgh, James Clerk Maxwell Building, Edinburgh EH9 3JZ, UK
⁴Institute of Molecular Cell and Systems Biology, University of Glasgow, Bower Building, Glasgow G12 8QQ, UK
⁵Department of Biological Sciences, University of Warwick, Coventry CV4 7AL, UK

Our understanding of the complex, transcriptional feedback loops in the circadian clock mechanism has depended upon quantitative, timeseries data from disparate sources. We measure clock gene RNA profiles in Arabidopsis thaliana seedlings, grown with or without exogenous sucrose, or in soil-grown plants and in wild-type and mutant backgrounds. The RNA profiles were strikingly robust across the experimental conditions, so current mathematical models are likely to be broadly applicable in leaf tissue. In addition to providing reference data, unexpected behaviours included co-expression of PRR9 and ELF4, and regulation of PRR5 by GI. Absolute RNA quantification revealed low levels of PRR9 transcripts (peak approx. 50 copies cell⁻¹) compared with other clock genes, and threefold higher levels of LHY RNA (more than 1500 copies cell⁻¹) than of its close relative CCA1. The data are disseminated from BioDare, an online repository for focused timeseries data, which is expected to benefit mechanistic modelling. One data subset successfully constrained clock gene expression in a complex model, using publicly available software on parallel computers, without expert tuning or programming. We outline the empirical and mathematical justification for data aggregation in understanding highly interconnected, dynamic networks such as the clock, and the observed design constraints on the resources required to make this approach widely accessible.

1. Introduction

Circadian clocks are found widely among organisms from archaea to mammals [1,2]. These internal time-keepers generate approximately 24 h rhythms in the expression of 10–30% of genes, even without environmental cues. In natural conditions, circadian rhythms are entrained by light and temperature cycles. Their function is to coordinate internal processes with the external day/night cycle [3,4] and also, through photoperiodism, relative to the seasonal cycle [5]. The circadian system of each organism includes a phylum-specific gene regulatory network that is required for most rhythmicity [6], as well as non-transcriptional oscillator(s) that are less well characterized in eukaryotes [7].

In plants, the clock gene network includes highly connected, negative regulators forming a complicated circuit. This has been best studied in Arabidopsis thaliana. One simplification (figure 1a) visualizes the circuit as a three-loop...
structure of morning and evening loops coupled around a repressilator [10,11]. The morning loop includes the MYB-related transcription factors LHY and CCA1, which activate expression of the pseudo-response regulators PRR9 and PRR7 [12,13], but inhibit expression of later-expressed genes including PRR5 and TOC1 (PRR1). PRR9, PRR7, PRR5 and TOC1 bind to and inhibit LHY and CCA1 expression, as predicted by modelling [10,14] and demonstrated by experiments [15–18]. LHY and CCA1 also inhibit expression of ELF3, ELF4 and LUX (PCL1), whose protein products interact to form another repressor, the evening complex (EC) [19–22]. The EC is thought to inhibit the expression of at least ELF4 and LUX, forming a negative feedback loop, whose continued function might explain the damped oscillation of clock gene expression observed in lhy cca1 double mutant plants [10].

GI, a large plant-specific protein, is also rhythmically expressed but functions at a post-translational level through, for example, stabilization of the TOC1-degradation factor ZTL [23]. Light signalling controls multiple processes (electronic supplementary material, figure S1) that entrain the clock circuit to the day–night cycle. A growing number of identified processes and components remain to be fully integrated into the circuit, though even the components described are challenging to analyse.

Figure 1. The clock gene network and experimental protocols. (a) The clock gene network summarized in the activity-flow language of SBGN v. 1.0 [8], with the principal connections in the P2012 model [9]. The repressilator is denoted by green lines; morning loop components are filled yellow; LHY/CCA1, red; evening loop components, blue. Light inputs are shown in electronic supplementary material, figure S1 and all modelled connections of P2011 [10] in electronic supplementary material, figure S2. (b) Peak-normalized RNA profiles of genes depicted in (a), in plants of the Col-0 accession under a 12 h light : 12 h dark cycle (LD 12 : 12; experiment 2b of panel (c)). (c) Graphical representation of the growth conditions. Experiments 1, 4, 5, 6 and 7 used seedlings grown in LD 12 : 12 for the number of days indicated; experiments 2 and 3 used plants grown on soil in LD 12 : 12 for the number of days indicated. Sucrose concentrations, growth temperatures and genotypes tested are shown for each experiment. Open box, light interval; black box, dark interval; light grey box, predicted darkness in constant light; dark grey box, predicted light in constant darkness; red box, red light. Sampling time in ZT (h), relative to lights-on of the first day of sampling or the last dawn before experimental treatment (ZT0). Ros, rosette; sd, seedling.
Formal, mathematical models have been helpful in understanding the plant clock, because its components are highly interconnected by nonlinear regulation (electronic supplementary material, figure S2; reviewed in [24]). Model development was necessarily based upon timeseries data, where the system was manipulated using mutations and by varying light or temperature inputs. More detailed models demanded greater precision and breadth in the data, which raised two major issues. First, data collation was laborious, because the numerical data underlying published timeseries graphs were rarely accessible [25]. Although the potential benefits of data sharing are recognized [26,27], in practice, useful sharing requires cyber infrastructure, which is currently best-developed for omics data rather than the many focused studies in the clock literature [28]. Second, the published data on Arabidopsis clocks used several genetic backgrounds and growth conditions, introducing ill-defined variation to the results.

To provide directly comparable data, we conducted large-scale qRT-PCR assays for the RNA levels of multiple clock genes. Overlapping studies in four laboratories using different growth stages and conditions highlighted the robustness of most expression profiles and the few instances where they varied. Visualizing the data as phase plane plots suggested new dynamic interactions and their genetic regulators. Absolute RNA quantification revealed the low expression levels of ELF3 and PRR9. To facilitate similar projects, we introduce data aggregation in the online BioDare resource, and illustrate the utility of our datasets by reoptimizing the P2011 clock model [10] with the open-source application SYSTEMS BIOLOGY SOFTWARE INFRASTRUCTURE (SBSI) [29], highlighting key areas for future experiments.

2. Results

2.1. Large-scale measurement of clock gene RNA profiles

This study was motivated by two projects that integrated circadian regulation into research on other plant physiological systems, which were incompatible with the growth conditions used in earlier circadian research. The Regulation of Biological Signalling by Temperature (ROBuST) project studied the interactions of ambient temperature with circadian and light signalling circuits; exogenous sucrose inhibits light signalling [30,31] and was therefore excluded. The Timing of Metabolism (TiMet) project studied circadian regulation of the starch pathway, among others, which is best characterized in rosette plants grown on soil. To measure the rhythmic expression in a set of clock-related genes (figure 1b), we used automated systems in Golm and Edinburgh to quantify mRNA levels for components of the clock circuit every 2 h, in multiple conditions and mutant backgrounds [32,33] (figure 1c). The ROBuST dataset tested 13-day-old, wild-type (WT) and mutant seedlings grown at 17°C on agar medium without additional sucrose. Datasets from the TiMet project tested 21-day-old rosette plants grown at 20°C on soil (TiMet ros) and 13-day-old seedlings on soil (TiMet sd1). The TiMet rosette data were collected from WT and clock mutant Arabidopsis thaliana plants grown under light: dark (LD) cycles in two experiments, followed by constant light (LL) or constant dark (DD) in one study. Three further studies were compared, from seedlings grown on sterile agar media without sucrose (TiMet sd2, using the same medium as the ROBuST data), or with exogenous sucrose under white (McWatters, this paper; and Edwards et al. [34]) or red light (Southern, this paper; and [21,35]).

2.2. Data presentation

Time is expressed as zeitgeber time (ZT) in hours since the last dark–light transition, by convention; the first dark–light transition within the sampling interval is 0 h on our plots. TiMet data are presented as absolute values [33], obtained by calibrating RNA extraction efficiency with heterologous control RNAs (electronic supplementary material, table S1) to calculate the number of copies of each RNA per gram fresh weight (gFW). Estimated cell numbers per gFW (see electronic supplementary material) were used to calculate RNA copies per cell. The other datasets are normalized relative to a control transcript (ACTIN7 for ROBuST; ACTIN2 for Edwards and Southern; B-TUBULIN4 for McWatters). ACTIN2 and GAPDH controls were also assayed with two amplicons each in the TiMet assays, for comparison among datasets. Data were replicated in biological duplicate or triplicate samples and in equivalent sampling on successive days (0–12 h and 24–36 h in the TiMet and Edwards datasets). Data are presented on linear scales to reflect the potential for protein synthesis and hence regulatory effects on downstream targets (in keeping with most of the literature; figures 2 and 3; electronic supplementary material, figure S5) and on logarithmic scales to reveal the full dynamic range of RNA expression, and hence the influence of multiple upstream regulators (figures 4–6; electronic supplementary material, figure S3 and S4). Further technical comparison among the studies is presented in the electronic supplementary material.

2.3. Similarity and specific variations of wild-type RNA profiles across datasets

Clock gene RNA expression profiles in WT plants of two accessions (Col and Ws-2) grown in LD are presented in figure 2; profiles were similar across the TiMet and ROBuST datasets, despite major differences in growth conditions. The morning clock components, LHY and LHY, peaked as expected at dawn (figure 2a, b), followed by PRR7 (ZT6; figure 2c, d), PRR5 and GI (ZT8; figure 2e–i). The expression of the evening components, ELF4, ELF5 and TOC1, peaked at ZT8–12 (figure 2e–i); peak expression of ELF5 was delayed by about 2 h in Col plants relative to Ws-2 in both datasets (figure 2g, h; replicated in LL data). ELF5 had a low-amplitude profile in both datasets, with lowest expression around ZT4.

The TiMet and ROBuST datasets differed at particular timepoints for PRR9, GI and TOC1. PRR9 expression was highest at ZT2–6 in both cases, with a clear peak at ZT2 in the ROBuST seedling data (consistent with many other reports from seedlings) but a broader profile in the TiMet data (figure 2g, h). After its major peak at ZT8–12, TOC1 expression is consistently observed (since [36]) to increase around ZT18, but the level of this night-time peak varied (figure 2e, f). The ROBuST data for seedlings showed a peak of GI expression at ZT2 (figure 2f); little induction is evident at ZT2 in the TiMet rosette data on a linear scale (figure 2e) though the logarithmic scale reveals the response (electronic supplementary material, figure S3e). The morning peak in
GI is likely to be an acute response to lights-on. Rapid sampling in the Southern data [35] and in a follow-up microarray study [10] suggested that induction is rapid but transient, and therefore sensitive to sampling time. Nevertheless, the data suggest that either the magnitude or kinetics of light responsiveness vary across the conditions tested. The difference in PRR9 profiles could reflect slower activation of PRR9 in the TiMet data, consistent with lower light responsiveness in rosettes than in seedlings or with faster repression of PRR9 in seedlings. The level of GI transcripts at ZT12 also varied from 4% to 40% of the peak level, with the lowest level in rosettes of Ws-2 (figures 2e,f and 3c). GI expression is light sensitive at this phase [37], so our results are consistent with variation in light responsiveness.

Sucrose modestly increases expression of the evening clock components TOC1 and GI [38], particularly in darkness [39], and can repress PRR7 with subsequent effects on CCA1 under low light [40], along with transcriptome-wide effects under LD cycles [41,42]. We therefore compared the expression profiles for CCA1, TOC1 and GI in plants grown without (ROBuST and TiMet data) or with exogenous sucrose (McWatters, Edwards and Southern datasets; figure 3). To facilitate comparison, TiMet data were normalized to control transcripts (two amplicons each in GAPDH and ACTIN2), as for the other studies. Each profile was normalized to its maximum. Expression profiles of CCA1 across the different timeseries matched closely despite the differences in accession and experimental protocols (figure 3a). The times of peak, mid-rising and mid-falling phases differed by at most 2 h (one sampling interval) among datasets. In the falling phase at ZT4, the profiles in McWatters, TiMet ros and TiMet sd2 data were delayed relative to the other data. The night-time expression of TOC1 at ZT18 varied from 20% to 60% of the main peak level (figure 3b), with high expression in ROBuST, Edwards and TiMet sd2 datasets. The expression of GI at ZT2 in the TiMet and Edwards seedling data was about 20% of the main peak level (figure 3c, also in Southern data [35]), intermediate between the levels in ROBuST and TiMet rosette data (discussed above). These features of the expression profiles showed no clear relationship with growth medium or developmental stage.
2.5. Regulation of clock genes under environmental and genetic manipulation

The TiMet project measured clock gene expression in LL and DD following LD entrainment, in seedlings of two WT and four clock mutant backgrounds (figure 5), revealing novel aspects of clock gene regulation as well as replicating regulation observed in many earlier, smaller studies. The results are discussed below with respect to the upstream regulators of each gene, rather than the effect on the gene’s downstream targets. The RNA data are therefore presented in semi-logarithmic plots that show regulator activity even at low RNA levels.

Comparing the three environmental conditions, peak RNA expression levels tended to fall in LL, consistent with the loss of dark-dependent regulation. The acute gene induction at the dark–light transition, faster degradation of PRR repressors in darkness and of the EC in the light are all expected to enhance rhythmic amplitude in LD. Expression levels of the clock RNAs were maintained in the first cycle in DD, except for the strongly light-regulated ELF4 [43,44]. Comparing the six genotypes, mutations that removed the repressors revealed the key connections in the clock circuit (figure 1a). The gi mutation, in contrast, had small or negligible effects on the timing and levels of expression except for PRR5, as noted below.

2.5.1. LHY and CCA1

Our results are consistent with PRR repressors controlling both the rising and falling phases of LHY and CCA1 expression at the transcriptional level [14,16–18,45]; several observations suggest that this activity is light-dependent. Both transcripts retain strikingly higher expression in the prr7;prr9 double mutant than in the WT, at ZT6–12 in LD and LL (figure 5a,b; p < 0.05; 20- to 30-fold higher at ZT8), consistent with the absence of repression from PRR9 and PRR7 proteins. By the second day in LL, the trough of LHY and CCA1 expression at ZT44 (68 h in figure 5) was also 20-fold higher than the WT trough level at ZT36–38 (60–62 h). Comparing LD and LL data with DD conditions revealed broader peaks of LHY and CCA1 RNA in DD (figure 5k,l), consistent with slower degradation of these transcripts in darkness [34,46]. In darkness, however, LHY and CCA1 levels in the prr7;prr9 mutant behaved very similarly to the WT, both during the falling phase in DD (ZT28–38; figure 5k,l) and during the rising phase in LD (ZT16–22; figure 5a,b). By dawn in LD, both transcripts peaked at the WT level, consistent with previous reports by trough levels) also varied greatly among clock genes. The TOC1 and ELF3 profiles showed only eight- to 20-fold amplitude in Col-0, and generally smaller amplitudes in other, mutant genotypes than the other clock genes (figure 4f–j), whereas LHY, CCA1, GI, ELF4 and PRR5 RNAs showed over 100-fold amplitude. This distinction was consistent in other datasets [21,34]. Amplitude estimates can be significantly affected by variation in the very low trough levels, which were higher in the TiMet sd1 dataset relative to the TiMet rosette data for LHY and all the evening-expressed genes in the Ws-2 accession, for example (figure 4). Transcripts with high-amplitude profiles might be expected to control circadian timing more effectively than the low-amplitude profiles of TOC1 and ELF3.

2.4. Absolute quantification of clock gene transcripts

The absolute quantification in the TiMet rosette data, which is based ultimately upon the certified amounts of synthetic commercial standards [33], revealed wide variation in peak RNA levels among clock genes in WT plants (figure 4). Highest RNA levels were detected for LHY at 1000–2100 copies per cell, similar to the control genes GAPDH and ACT2. PRR9 was least abundant at the peak, with 40–70 copies per cell; LUX and ELF3 peaked at 105–130 copies per cell; PRR7, PRR5, GI and TOC1 at 120–270 copies per cell; ELF4 and CCA1 at 250–600 copies per cell. RNA copy number of LHY was threefold greater than that of CCA1 (figure 4a,b).

Peak levels for the evening-expressed genes (figure 4f–j) were slightly higher in Ws-2 than Col-0 plants, by 1.2-fold (LUX) to 2.0-fold (ELF4), average 1.6-fold. Several clock gene RNAs fell to low copy numbers per cell at the trough. Consequently, rhythmic amplitudes (defined here as peak divided
Figure 4. Range of transcript abundance for clock genes in clock mutants. The bars show the highest and lowest mean values for the absolute abundance of transcripts for clock genes in a given genotype. The genotypes are, from left to right, Col-0 wild-type, gl-201, prr9 prr7 double mutant, toc1, WS WT, lhy cca1 double mutant (from experiments 2 and 2B of figure 1c, 21-day-old rosettes) and WS (designated WS_2) and elf3 from experiment 3 (13-day-old seedlings), (a) LHY, (b) CCA1, (c) PRR9, (d) PRR7, (e) PRR5, (f) TOC1, (g) LUX, (h) GI, (i) ELF3, (j) ELF4. The underlying data are as in figures 5 and 6.
Figure 5. Clock gene expression in wild-type plants and clock mutants in LD, and after transition to constant light (LL) or darkness (DD). Col-0 and Ws-2 WT, the lhy-21 cca1-11 and prr7-3 prr9-1 double mutants, and the toc1-101 and gi-201 single mutants were grown in a 12 h photoperiod for 20 days, harvested through a LD cycle and then transferred to LL (a–j) or DD (k–t); TiMet ros, dataset 2 of figure 1c). Transcript levels for clock genes were measured by qRT-PCR, including eight external RNA standards to allow absolute quantification. (a,k) LHY, (b,l) CCA1, (c,m) PRR9, (d,n) PRR7, (e,o) PRR5, (f,p) TOC1, (g,q) LUX, (h,r) GI, (i,s) ELF3, (j,t) ELF4. The results are the mean of duplicate samples, error bars show SD. Open box, light interval; black box, dark interval; light grey box, predicted darkness in LL; dark grey box, predicted light in DD.
Figure 5. (Continued.)
Figure 6. Clock gene expression in wild-type plants and elf3 mutants in LD. Ws-2 WT (solid lines) and elf3 – 4 mutant plants (dashed lines) were grown in a 12 h photoperiod for 12 days and harvested through one LD cycle (TiMet sd, dataset 3 of figure 1c). Transcript levels for clock genes were measured by qRT-PCR, including eight external RNA standards to allow absolute quantification. (a) LHY, (b) CCA1, (c) PRR9, (d) PRR7, (e), PRR5, (f), TOC1, (g) LUX, (h) GI, (i) ELF3, (j) ELF4. The results are the mean of duplicate samples. Error bars show SD.
[12,13]. Thus, the misregulation of LHY and CCA1 in the light in the prr7;prr9 double mutant was abolished during the dark in LD.

Removing TOC1, the last of the PRR repressors to be expressed, would be expected to allow an earlier rise in expression of LHY and CCA1 during the night in the toc1 mutant under LD. This effect was relatively small (two- to 2.5-fold higher at ZT18, \( p = 0.02 \)). LHY and CCA1 levels in toc1 mutants differed less than fourfold from WT at any point in LD. The mutant phenotype was not enhanced in the first DD cycle (figure 5k,l). In LL, however, LHY and CCA1 expression in the toc1 mutant peaked at ZT22 (46 h in figure 5) rather than at ZT26 (50 h) in Col, reached only 30–50% of WT peak level consistent with earlier data [47], and fell much earlier than the WT (19– to 27-fold lower at ZT30, time 54 h in figure 5a,b). Thus, the molecular phenotypes of both prr7;prr9 and toc1 mutants were light-dependent.

The elf3 mutant reduced peak expression of both LHY and CCA1 by five- to 10-fold (figure 6a,b; electronic supplementary material, figure S5g), with greatest effects at ZT20–24. This effect is thought to be indirect, as the EC (comprising ELF4, ELF3 and LUX) is proposed to repress the PRRs (figure 1a), as well as LUX and ELF4 [10,19,20,22]. De-repression of PRR expression in mutants of the EC should therefore explain the effects of elf3 on LHY and CCA1.

### 2.5.2. PRR9 and PRR7

PRR7 was the most severely affected gene in the elf3 mutant under LD, maintaining 25–85% of the WT peak level at all times (figure 6d), consistent with de-repression of the PRR7 promoter [21]. The resulting, 30- to 50-fold overexpression of PRR7 in elf3 at ZT20–24 is consistent with reduced expression of LHY and CCA1 at this time. PRR9 transcript levels retained a 100-fold rhythmic amplitude under LD in the elf3 mutant, indicative of ELF3-independent regulation (see Discussion). Nonetheless, PRR9 expression was also de-repressed from ZT10 in elf3 (\( p = 0.05 \)), rising 2–4 h before dawn (figure 6c), and presumably also contributing to reduce LHY and CCA1 expression. The early-expressed PRRs are thought to be repressed by the later-expressed PRR5 and TOC1 (figure 1a). The toc1 mutation had modest effects on PRR9 or PRR7 profiles under LD cycles (figure 5c,d), though the changes observed (such as an early rise in PRR7 at ZT20–24) were not consistently significant in the TiMet and ROBuST datasets, or in DD in the TiMet data (figure 5n). toc1 also had little effect on LHY and CCA1 levels in these conditions (figure 5a,b). In LL, however, removing TOC1 prevented full repression of the PRRs. The trough of PRR7 expression was at a 10-fold higher level than in the WT (\( p < 0.05 \)) and 8 h earlier (ZT12 rather than ZT20, 36 h rather than 44 h in figure 5d). Higher expression of the repressor PRR7 at 38–52 h (figure 5d) is consistent with the lower peak expression of CCA1 and LHY in toc1 under LL (figure 1a [9]). Taken together, these results suggested that TOC1 repressor function was most effective under constant light conditions, where the toc1 mutant was originally identified [48]. Light-dependent regulation was also evident in WT plants. Peak PRR9 expression levels fell less than twofold in the first cycle of DD (\( p > 0.16 \); figure 5m). Peak PRR7 expression tended to increase (threecold or less) in all genotypes in DD (figure 5n; electronic supplementary material, figure S4c).

The gi mutant was an exception, which slowed the rise of all the transcripts in DD except ELF3 and ELF4 (figure 5k–l; electronic supplementary material, figure S5d). Peak expression for some genes was reduced in gi below WT levels, including PRR7 (\( p = 0.02–0.03 \) at ZT26–28 h). Trough RNA levels in the WT plants rose more dramatically in DD, for PRR7 and other clock genes (except for LHY): the lowest expression of PRR7 in Col was 1.5 ± 0.4 copies per cell at ZT20 but 65 ± 6.8 copies per cell at ZT40 (electronic supplementary material, figure S4c). The Edwards dataset showed similar de-repression of CCA1 and GI trough levels in DD (electronic supplementary material, figure S4a,b) [34]. Lastly, we tested the effect of CCA1 and LHY on the PRR transcripts, using the lhy;cca1 double mutant. In the WT, the repression of the evening-expressed genes by LHY and CCA1 in the early day delays the expression of these and other target genes until the evening. The double mutation advanced the peak phase of all the other clock genes to ZT2–4, as expected, except for ELF3 (see below). Despite the de-repression, peak levels were not consistently increased relative to the WT-2 control. Peak expression of PRR9, PRR7 and PRR5 (figure 5c–e) was slightly reduced (up to twofold) in the lhy;cca1 background under LD, consistent with earlier results [12]. By ZT8 (or ZT4 for PRR9), all the clock genes were expressed at lower levels in lhy;cca1 than in the WT (\( p < 0.01–0.04 \)), consistent with expression of all the PRR repressors. In the lhy;cca1 double mutant in DD, however, the PRR genes had broad peaks that rose earlier than in the WT (ZT22–30) but did not fall earlier (ZT34–40; figure 5m–o; electronic supplementary material, figure S4d). The absence of early repression in DD again suggests that inter-regulation of the PRRs is light-dependent.

### 2.5.3. PRR5 and TOC1

The later-expressed PRRs are repressed by LHY and CCA1, so longer expression of LHY and CCA1 in the prr7;prr9 double mutants delayed their expression in LD and LL conditions (figure 5e,f), as expected. In contrast, under DD conditions, PRR5 expression in prr7;prr9 rose indistinguishably from the WT at ZT26–34 h and peaked slightly (twofold) above the WT level (figure 5o). The lhy;cca1 double mutant caused the phase advance noted above, as the loss of LHY and CCA1 repressors increased TOC1 levels in the early day. Peak TOC1 RNA levels in the lhy;cca1 mutant did not change consistently from WT levels in the TiMet data under LL (figure 5f), and were lower than the WT in the ROBuST dataset (\( p < 0.01 \); electronic supplementary material, figure S5c).

Our detailed datasets also allowed us to compare expression waveforms. For example, PRR5 rises and falls 10-fold within 5 h in both TiMet and ROBuST data (figure 5e,o; electronic supplementary material, figures S3h, j, S5a,b). This narrow peak indicates highly nonlinear control, consistent with negative autoregulation and/or inhibition by TOC1 [15,49]. Moreover, our results indicate that this PRR5 waveform depends upon GI function. The gi-201 mutant had limited effects overall but slowed the fall in PRR5 mRNA in LD and LL (figure 5e), creating an asymmetric profile in PRR5 RNA that was also observed in the gi-11 tested in the ROBuST data (electronic supplementary material, figure S5b). Repression by the EC might also contribute to the falling phase of the PRR5 profile. Removing this repression in the elf3 mutant resulted in moderate de-repression of PRR5 and TOC1 in the...
late night (p < 0.01, ZT0/24; p = 0.01 for PRR5 ZT22) and potentially in the early morning (p = 0.06–0.08; ZT2–4; figure 6c,f). In contrast, de-repression of the early PRRs in elf3 was greatest in the early night (see above), indicating that the profile of regulators varies among the PRR family members (see Discussion).

2.5.4. GI

The main peak of GI expression in the late day behaves similarly to PRR5, with delayed expression in the prr7;prr9 double mutant owing to longer expression of LHY and CCA1 under LD and LL but not DD, and an advanced phase in the lhy;cca1 double mutant (figure 5h,p). In contrast to PRR5 but similarly to PRR9 and PRR7, GI was de-repressed from ZT10 in the elf3 mutant (p < 0.01), consistent with [21] and the Southern dataset (electronic supplementary material, figure S5f). The Southern dataset showed that the expression of GI was similar in elf3 and elf4 mutants, but there was much less effect on CCA1 in elf4 than elf3 (electronic supplementary material, figure S5g), indicating that the effects of the EC components can be distinct.

2.5.5. ELF3

The ELF3 rhythmic profile has low amplitude, as noted above, with a trough at ZT2–4 and peaks at both ZT8 and ZT18–20 in WT plants under LD in the TiMet and RObuST datasets (figure S5i,s; electronic supplementary material, figures S3–j, S5d). The trough of ELF3 expression is de-repressed at ZT4 in the lhy;cca1 double mutant (p < 0.01), though there is no peak at this time, in contrast to all the other clock genes. The rise in ELF3 expression is delayed in the prr7;prr9 double mutant (p < 0.01–0.05, at ZT6–10), consistent with repression by increased levels of LHY and CCA1 (figure 5i). The elf3–4 allele contains a small deletion in the coding region [50] resulting in negative autoregulation (electronic supplementary material, figure S5g). This result is consistent with LUX binding to its cognate promoter [20] resulting in negative autoregulation (figure 1a [10]). ELF4 expression in the elf3 mutant, in contrast, showed a pattern more similar to TOC1 and PRR5 (see above), with de-repression only from ZT22–ZT6 h (figure 6i).

2.6. Alternative visualization gives new insights into co-regulation of clock genes

Data visualization is critical in analysing the complex interactions within the clock gene circuit, in order to generate new hypotheses. Timelines plots do not show these interactions directly. They can be revealed in phase plane diagrams that plot the levels of two components against each other (figure 7), though this format is less familiar (see electronic supplementary material). First, phase plane plots emphasize the relative timing of clock components, rather than control by the light : dark cycle. For example, GI rose without (before) TOC1, especially in Col plants of the TiMet and RObuST datasets that were grown without exogenous sucrose. High TOC1 levels extended later than high GI, particularly in Ws-2 plants of the TiMet datasets (figure 7a). Second, this visualization can reveal interactions among the components plotted. For example, figure 7b shows TOC1 RNA levels in younger plants were maintained at 35–55% of the peak level at ZT20–22, when CCA1 expression rose above 50% of its peak level. TOC1 levels were lower for the same CCA1 level in rosette plants. The logarithmic scale shows this more clearly (figure 7c). This suggests that CCA1 protein is not yet an effective repressor of TOC1 at this phase, especially in younger tissues.

Finally, the phase plane diagrams can show how the interaction of two genes depends upon a third regulator. Expression peaks of PRR9 and ELF4 were far out of phase in the WT (figure 7d), for example. Data from LL (filled symbols) suggest a negative correlation in the subjective night, when ELF4 falls as PRR9 rises. However, the two genes peak then fall together in the lhy;cca1 double mutant under LL, at ZT26 and ZT42 (figure 7e; equivalent to timepoints 50 and 66 h in figure 5), creating a diagonal with a positive gradient (red dashed line, figure 7f). PRR9 also had an earlier peak that was not shared by ELF4 (ZT22 and ZT38, or 46 and 62 h in figure 5; black arrowheads in figure 7f). Both features were reproduced on two successive cycles, though PRR9 expression was less than 1% of the WT peak level. Thus LHY, CCA1 and the LD cycle all differentiate PRR9 expression from ELF4, but in their absence, PRR9 and ELF4 expression profiles are similar for much of the circadian cycle (six of eight timepoints in the short, 16 h cycle of the mutant), presumably controlled by the other PRRs and/or the EC. Likewise, phase plane diagrams for the prr7;prr9 double mutant (electronic supplementary material, figure S6) suggested that not only CCA1 and LHY, but also the PRRs repress ELF4 in the WT. In addition to visualization, many other aspects of data management benefit significantly from online data infrastructure.

2.7. Online infrastructure for data sharing

Our open-source BioDare (Biological Data repository) [51] supports data from many small-scale experiments that collectively represent a significant resource (table 1). Empirical evidence indicates that these data are essential to understand complex biological regulation, and mathematical analysis shows why
this is the case (see Discussion). In addition to six rhythm-analysis algorithms [52] and protocols for analysis, statistical summary and visualization [53], BioDare facilitates data sharing and public dissemination by providing a stable identifier for each experiment. Detailed metadata (experimental description) ensure that the data can be reused appropriately. Results can be compared across studies and laboratories (‘data aggregation’) by searching the metadata for genotype, marker gene and other terms (figure 8). Increased expression of GI in the elf3 mutant, for example, is highlighted despite the greater technical variability of manual assay preparation in the Southern dataset compared with the later, robotized assays in the TiMet data (figure 6; electronic supplementary material, figure S5f and Methods).
2.8. Optimizing clock models with public resources

One goal of such comparisons is to determine how much of the available data is matched by a particular mathematical model: the ROBuST and TiMet experiments were designed to test models of the clock gene circuit under different growth conditions. However, testing complex models against large datasets requires skills that are rare among plant molecular researchers. We therefore tested whether our comprehensive data and better computational resources could make modelling more accessible. The open-source SBSI allows non-programmers to optimize model parameters in order to match diverse data, on large, parallel computers [29]. As a test case, we addressed a recognized limitation of the original P2011 model [10], termed P2011.1.1. The model was developed to understand circadian clock function under light–dark cycles and, separately, under constant light. Following a transition from LD to LL (as in figure 5a–j), the first peak in expression of the combined LHY/CCA1 component under constant light occurred at ZT28.4 h (52.4 h in figure 7a), about 2.5 h later than in the TiMet ros data (as noted [25,54]). The model’s light–dark function was replaced with the input signal step function [55] to represent the LD–LL transition in the community model exchange format, SBML [56]. The resulting model P2011.2.1 was optimized in SBSI (see electronic supplementary material), testing model simulations with many alternative parameter sets against the TiMet ros RNA dataset, including circadian period values for clock mutants and WT plants [29].

The optimized parameter set of model P2011.2.1 more closely matched the data, including an earlier peak of LHY/CCA1 in LL at ZT26.5 h (figure 9a) and close match to TOC1 and GI profiles in LD (ZT10–12 h; figure 9b,c), while retaining other qualitative behaviours. LHY/CCA1 expression rises in LL after the PRR repressor proteins are degraded. Consistent with this notion, removing TOC1, the last gene in the PRR repressor wave, advanced the phase of the entire clock mechanism in LL. Results for PRR7 are shown in figure 9d,e. PRR protein degradation rates were not strongly affected in P2011.2.1; rather, overall PRR levels were lower than in P2011.1.2 (not shown).

In the simulated toc1 mutant, the peak of LHY/CCA1 was 1.4 h earlier than simulated WT in P2011.1.2, 2.5 h earlier in P2011.2.1, but 4 h earlier in the data (figure 5a,b). The simulations of PRR7 show the same improved timing of the new function [55] to represent the LD–LL transition in the community model exchange format, SBML [56]. The resulting model P2011.1.2 was optimized in SBSI (see electronic supplementary material), testing model simulations with many alternative parameter sets against the TiMet ros RNA dataset, including the LD–LL transition (figure 5a–j), and against circadian period values for clock mutants and WT plants [29].

The optimized parameter set of model P2011.2.1 more closely matched the data, including an earlier peak of LHY/CCA1 in LL at ZT28.4 h (52.4 h in figure 7a), about 2.5 h later than in the TiMet ros data (as noted [25,54]). The model’s light–dark function was replaced with the input signal step function [55] to represent the LD–LL transition in the community model exchange format, SBML [56]. The resulting model P2011.1.2 was optimized in SBSI (see electronic supplementary material), testing model simulations with many alternative parameter sets against the TiMet ros RNA dataset, including the LD–LL transition (figure 5a–j), and against circadian period values for clock mutants and WT plants [29].

<table>
<thead>
<tr>
<th>research group</th>
<th>location</th>
<th>experiments</th>
<th>% total experiments</th>
<th>timeseries</th>
<th>% total timeseries</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. J. Millar</td>
<td>Edinburgh, UK</td>
<td>332</td>
<td>14</td>
<td>41 890</td>
<td>18</td>
</tr>
<tr>
<td>A. Hall</td>
<td>Liverpool, UK</td>
<td>261</td>
<td>11</td>
<td>79 228</td>
<td>34</td>
</tr>
<tr>
<td>D. Bell-Pedersen</td>
<td>Texas A&amp;M, USA</td>
<td>138</td>
<td>6</td>
<td>1428</td>
<td>1</td>
</tr>
<tr>
<td>J. Agren</td>
<td>Uppsala, Sweden</td>
<td>18</td>
<td>1</td>
<td>9370</td>
<td>4</td>
</tr>
<tr>
<td>K. J. Halliday</td>
<td>Edinburgh, UK</td>
<td>230</td>
<td>10</td>
<td>5043</td>
<td>2</td>
</tr>
<tr>
<td>L. Larrondo</td>
<td>Santiago, Chile</td>
<td>75</td>
<td>3</td>
<td>6429</td>
<td>3</td>
</tr>
<tr>
<td>M. Jones</td>
<td>Essex, UK</td>
<td>89</td>
<td>4</td>
<td>3148</td>
<td>1</td>
</tr>
<tr>
<td>M. Hastings</td>
<td>MRC LMB, UK</td>
<td>1071</td>
<td>45</td>
<td>58 770</td>
<td>25</td>
</tr>
<tr>
<td>S. Harmer</td>
<td>UC Davis, USA</td>
<td>37</td>
<td>2</td>
<td>11 353</td>
<td>5</td>
</tr>
<tr>
<td>S. A. Kay</td>
<td>USC, USA</td>
<td>38</td>
<td>2</td>
<td>12 972</td>
<td>6</td>
</tr>
<tr>
<td>All BioDare</td>
<td></td>
<td>2344</td>
<td></td>
<td>232 844</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Usage statistics of BioDare (Feb 2015), from originating groups and selected external users. An experiment represents a dataset similar to one of the above-described studies, which includes multiple timeseries, from samples of multiple genotypes, assays or reporters and/or environmental conditions. Totals include minor users that are not listed individually; the total number of data points is over 41 million.
Figure 9. Model re-optimization. Comparison of measured transcript levels from figure 5 (experimental data, symbols), with simulation of models P2011.1.2 (old model, dotted line) and P2011.2.1 (new model, solid line), which resulted from fitting to these data using SBSI. 0–24 h, LD; 24–72 h, LL. (a) LHY and CCA1 transcripts are combined in the model, so the average of LHY and CCA1 data is plotted. The peak of LHY/CCA1 under LL was delayed in the P2011.1.2 model (52.4 h) relative to the peak in the data (50 h), which was closely matched by the P2011.2.1 model (50.5 h). (b) GI transcript, (c) TOC1 transcript and (d) PRR7 transcript in Col-0 WT. (e) PRR7 transcript in the toc1 mutant shows a greater phase-advance in LL than either model. Chi-square cost value for match to TiMet ros Col-0 data in LD-LL was 20.2 for P2011.1.2, 7.6 for P2011.2.1. Chi-square cost for match to TiMet ros toc1 data in LD-LL was 39.7 for P2011.1.2, 13.1 for P2011.2.1.

Table 2. Optimization of model parameters from loose constraints. The starting P2011.1.2 model was optimized in SBSI to fit the TiMet ros dataset and additional period constraints (see electronic supplementary material, Methods). Model, version number of the resulting model. PlaSMo ID, model identifier in the PlaSMo resource. Job, computational job code. Start, the default parameters values from P2011.1.2 or nominal values (Nom). Range, the range of parameter values that were searched, either as fold change above and below the P2011.1.2 values or as a fixed range. Set-up trials, the number of randomly chosen parameter sets tested to initialize the optimization. Cost, the best cost value (closest fit to all constraints).

<table>
<thead>
<tr>
<th>model</th>
<th>PlaSMo ID</th>
<th>internal job ID</th>
<th>start</th>
<th>range</th>
<th>set-up trials</th>
<th>cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>P2011.1.2</td>
<td>PLM_71 ver1</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>171</td>
<td></td>
</tr>
<tr>
<td>P2011.2.1</td>
<td>PLM_71 ver2</td>
<td>.599</td>
<td>P2011.1.2</td>
<td>2–3 ×</td>
<td>5000</td>
<td>77</td>
</tr>
<tr>
<td>P2011.3.1</td>
<td>PLM_1041 ver1</td>
<td>t30</td>
<td>P2011.1.2</td>
<td>100 ×</td>
<td>2 097 152</td>
<td>175</td>
</tr>
<tr>
<td>P2011.4.1</td>
<td>PLM_1042 ver1</td>
<td>t34</td>
<td>Nom</td>
<td>0.001–10</td>
<td>67 108 864</td>
<td>270</td>
</tr>
<tr>
<td>P2011.5.1</td>
<td>PLM_1043 ver1</td>
<td>t37</td>
<td>Nom</td>
<td>0.001–10</td>
<td>67 108 864</td>
<td>190</td>
</tr>
<tr>
<td>P2011.6.1</td>
<td>PLM_1044 ver1</td>
<td>t40</td>
<td>Nom</td>
<td>0.0005–20</td>
<td>134 217 728</td>
<td>185</td>
</tr>
</tbody>
</table>

model for the WT (figure 9d) and the toc1 mutant in LD (figure 9e), but an earlier phase of the toc1 mutant data under LL. Regulatory interactions among the PRR genes will repay further analysis [9,10] in future models (see Discussion).

The computation time required for P2011.2.1 was only approximately 30 core-hours, because the model parameters were varied within only a narrow range (two- to threefold change) from their starting values in P2011.1.2 [10]. The P2011.2.1 parameters had been manually determined to match a wide range of data and qualitative behaviours in the clock literature; many were derived from the parent model P2010 [14]. When the first model of a system is developed, in contrast, most or all parameter values may be unknown. We therefore tested our approach in such scenarios (table 2). Allowing parameter values to differ by up to 100-fold from the values in P2011.1.2 created a very large parameter space.
that was nonetheless centred on a known, viable region. In contrast, starting parameters from nominal values (0.1, 1, etc.) and testing each parameter over the same range (such as 0.001–10) removed that anchor. Viable parameter sets that gave cost values similar to the unmodified P2011.2.1 were identified in each test, with computation times up to four core years for P2011.6.1, using the UK national supercomputing resource HECToR. These parameter sets are not intended to replace P2011.2 but to demonstrate that similar results can be achieved by a more accessible approach using the TiMet data and SBSI, without new programming or laborious, manual model development. The P2011 model versions and the cognate graphical network diagram (electronic supplementary material, figure S2) are publicly accessible from the PlaSMo repository and elsewhere (see appendix A).

3. Discussion

3.1. Robust regulation of clock gene expression

Quantitative timeseries data are crucial to understand the dynamics of any moderately complex regulatory system. As understanding advances, more precise questions can be formulated that demand both consistent and comprehensive datasets. We provide such data for the RNA profiles of genes associated with the Arabidopsis circadian clock, with an online resource to facilitate comparisons within and across datasets. Our experiments were designed to test clock function under the distinct conditions required for separate studies, on light signalling (in the ROBuST project) and carbon metabolism (in the TiMet project), using different technical platforms. The results presumably include the variation previously observed among experiments designed to be replicated across laboratories [57]. We compared two Arabidopsis accessions. Significant differences in circadian timing have been demonstrated among Arabidopsis accessions, albeit using long-term, imaging assays that integrate the effects of small timing changes over many cycles [58–60]. Importantly, the rhythmic RNA profiles tested here were remarkably consistent (figure 3). Progress in understanding the clock gene network must, in part, be attributed to this robustness of circadian regulation.

Several clock genes are regulated with high daily amplitude, more than 100-fold for LHY, CCA1, GI, ELF4 and PRR5 under LD (figure 4; electronic supplementary material, figures S3 and S4 [21,34]), falling to low RNA copy numbers per cell. Our data necessarily reflect the mean expression across cells in the rosette, greater than 80% of which are in the leaf mesophyll [61]. Nonetheless, the absolute calibration of our RNA assays provides one approach for future estimation of the average copy number for the cognate proteins.

The most striking variations of RNA profiles among WT plants involved the acutely light-responsive genes GI and PRR9. The ROBuST dataset showed the highest levels of GI and strong induction of PRR9 at ZT2 (figures 2 and 3). This is consistent with strong light induction, which might be mediated by direct photoreceptor signalling and/or by indirect sugar signalling. The absence of exogenous sucrose in the ROBuST conditions was not the sole cause, as the TiMet sd2 data used the same, sucrose-free media but did not show such strong GI induction (figure 3c). The lower growth temperature in ROBuST conditions (17°C rather than 20–22°C in other datasets) might also increase light responsiveness.

Consistent with this notion, both exogenous sucrose and higher ambient temperature limit other light responses [30,31].

3.2. Regulation of the PRR repressors

RNA profiles of the PRR gene family varied among datasets in the WT under LD, as well as among conditions and genotypes. The variable expression of TOC1 around ZT18 (figure 3b) awaits a mechanistic explanation, as do the depression of multiple genes in DD (for example, figure 5n) and of PRR5 in the gi mutant (figure 5e; electronic supplementary material, figure S5b). TOC1 is thought to be an active repressor at ZT18, so variable auto-repression is possible and might also explain variation in PRR5 expression at this phase (figure 2g). Alternatively, TOC1 expression might rise during a transition between one repressor in the early night (such as the EC) and another in the late night (such as LHY and CCA1).

The tight interconnections among the clock components complicate the analysis of these data, though the resulting combination of direct and indirect effects is now interpretable. For example, removing EC regulation in the elf3 mutant de-repressed the direct EC targets PRR9 and PRR7 in the early night, when the EC is active in WT plants. PRR5 and TOC1 were noted as potential targets based on mutant RNA profiles [10], but both genes were de-repressed around dawn in elf3, suggesting that an indirect mechanism owing to lower LHY and CCA1 levels is more significant than the loss of direct regulation by the EC in the mutant. PRR9 and PRR7 are both proposed EC targets (along with ELF4 and LUX), yet PRR9 (and ELF4) retains rhythmic regulation in the elf3 mutant under LD, whereas PRR7 (and LUX) is more severely affected (figure 6). To understand such differences in response, it will now be important to measure the affinity of regulators for their target genes, extending initial data [62]. Previous modelling results indicated that the different daily profiles of the PRR genes allow flexible responses to dawn and dusk [14], so the mechanisms that generate the PRR profiles will repay further analysis [10,11].

Several results suggested that regulation by the PRR genes is light-dependent. First, in the prr7;prr9 double mutant, LHY and CCA1 expression was de-repressed during the day but returned to match the WT profile at night in LD (figure 5). One explanation might be that PRR9 and PRR7 (directly or indirectly) antagonize the light activation of LHY and CCA1 during the day in the WT [14,63], and the absence of these PRR proteins in the double mutant has little effect in darkness. Consistent with this notion, the prr7 single mutant also showed a day-time de-repression of CCA1 in the ROBuST dataset (electronic supplementary material, figure S5e), albeit less than in the double mutant. However, the CCA1 profile in the prr7 single mutant was unaffected in the daytime, but de-repressed 2 h earlier in the night (electronic supplementary material, figure S5e). Thus, inter-regulation of the early PRR genes is important, in addition to regulation by TOC1 [10]. Second, in the lhycca1 double mutant, PRR gene expression is repressed to low levels at the end of the day in LD, consistent with simultaneous, early expression of all the PRR repressors in these mutant plants. In DD, however, the falling phase of PRR expression is the same in WT and double mutant plants (figure 5). The higher and earlier expression of the PRR RNAs in the double mutant in DD does not appear to be effective in
suppressing PRR expression. The faster degradation of the PRR proteins in darkness presumably contributes to these effects; it will be interesting to determine whether the interaction of the photoreceptor PHYB with clock proteins (including TOC1 [64]) also mediates the light sensitivity of this process.

3.3. Effects of exogenous sucrose

Current models of the Arabidopsis circadian clock are necessarily based on disparate data, much of it derived from seedlings grown on media containing high levels of sucrose. The presence or absence of exogenous sucrose under the conditions tested here affected the clock RNA profiles less, or at least no more, than other experimental parameters, despite the widespread regulation of plant genes by sucrose [41,65]. Consistent with this, effects of exogenous sucrose on clock gene expression in WT plants have previously been reported under very low light fluence rates or in the presence of photosynthetic inhibitors [40], in DD, CO2-free air or the starchless pgm mutant [39,66,67]. PRR7 was induced in sugar-starved conditions (extended DD and at night in pgm) and was repressed by resupply of 3% exogenous sucrose (electronic supplementary material, figure S7a) [67]. The TiMet rosette study tested PRR7 in DD, finding increased PRR7 levels (figure 5a), especially in the trough of the profile (electronic supplementary material, figure S4c). Trough levels of CCA1 and GI were also raised in DD in the TiMet data, and in the Edwards experiment that included 3% exogenous sucrose (electronic supplementary material, figure S3) [34]. De-repression of the trough levels in DD is neither specific to PRR7 nor to sugar limitation. Transcript levels of the TOC1- and PRR5-degrading F-box protein ZTL, and its homologues LKP2 and FKF1, also rose slightly in sugar-starved conditions (electronic supplementary material, figure S7b [67]), suggesting one possible mechanism for de-repression of PRR7 via faster degradation of PRR repressors.

3.4. Open resources for small-scale results

Our results will be useful to generate and test many hypotheses beyond those reported here. The potential for such future value might, in principle, justify the additional effort in curating and disseminating our data. In practice, future value motivated little data sharing, compared with present value. We therefore outline the mathematical understanding of and empirical evidence for such present value, together with practical steps that increased both present and future value relative to the effort involved in sharing data.

No suitable community repository existed for our results. One reason was the relatively large effort required to describe accurately many small data files, which deters researchers and resource developers from sharing such data [68]. The largest-scale omics and sequencing studies have different data structures, motivations, stakeholders and economics, which can facilitate data sharing [69] including exemplary resources in the circadian field [70–72]. However, mathematical analysis explains why the results of small-scale experiments are often particularly valuable in understanding biological systems. Gutenkunst et al. [73] showed that parameters were ‘sloppy’ in dynamic models of a range of biological regulatory systems, meaning that a wide range of parameter values could generate the simple behaviours that they tested. Rand et al. [74–76] tested how many parameter changes could affect the dynamic behaviour of such systems. All possible behaviours were tested and only a handful of behaviours could be readily achieved by changing parameters (these behaviours have also been termed the ‘dynatypic’ of the system, by analogy to the phenotype of an organism [77]). For circadian clocks, a change in period was the most accessible behaviour: many different parameter changes altered period under constant conditions [74]. The related, empirical result is that genetic screens seeking mutants with altered circadian period have not only identified clock components, but also many genes that affect the clock less directly [78,79]. Observing a change in period gives little evidence for the role of the mutated gene in the plant and does not strongly constrain any particular parameter in the model, but rather has a small constraining effect upon a large number of parameters, in agreement with Gutenkunst et al. [73]. A measured period value can therefore easily be accommodated without fundamentally changing the model. In contrast, manipulating the system to test less accessible behaviours provides strong constraints, albeit potentially on fewer parameters [76]. It is much more likely that such results would not be accommodated by any reasonable parameter values, falsifying the current model and leading to new understanding during the development of a better model. Thus, the number of manipulations tested is crucial; model analysis can prioritize the most informative manipulations [80,81].

One consequence for experimental design is that the number of manipulations is more important than the number of components tested. This concept is familiar from the statistical clustering of microarray timeseries. The behaviour of a single cluster mean can adequately represent hundreds of individual transcripts, even for genes with complex light and circadian regulation [82]. The individual transcript data are more valuable in identifying coregulated, downstream genes than in understanding the clock system upstream. Thus, targeted qRT-PCR or reporter gene assays have been more widely used in understanding the clock gene circuit, although they lacked a data-sharing resource. Despite the limited justification for costly omic assays, targeted data-sharing resources [67,70] have ensured that a subset of transcriptomics data have been reused effectively in clock studies.

Empirical evidence for the value of multiple manipulations comes from 10 years of modelling the plant clock gene circuit and output pathways. Constraining the models with timeseries data from many conditions was a critical tool [83], resulting in multiple, experimentally validated predictions. Gleaning the data from electronic supplementary material or by ‘scrapping’ numerical values from published charts made this possible. In practice, aggregating the numerical data has often taken a major effort, after which the data were shared on author web sites [25,84] or on BioDare [82,85,86].

BioDare [51] was developed to share timeseries data from relatively small-scale experiments conducted within individual laboratories (such as the Edwards, Southern and McWatters datasets) or in collaborative projects with few partners (such as ROBuST and TiMet). A regular user might upload an experiment with several hundred timeseries each week [87,88]. However, the user must also provide experimental metadata that are sufficiently detailed to pinpoint the most relevant experiment among hundreds to thousands of similar studies (table 1). The resource must therefore streamline the process of writing the structured metadata to minimize the weekly effort involved, and then use the metadata to provide
and their slow degradation delays the rise in slower degradation of PRR proteins in the light than in the model’s 2 h late phase in LL (figure 9) is caused by the constrained (also noted in references [54,88]). The P2011.2.1 cadian clock models are therefore equally applicable to the about such circuits, so long as the starting model adequately identifying investment in infrastructure to support their analysis. Mathematical models can powerfully express hypotheses that could be linked to organism-specific databases [91]. BioDare complements our repository of plant systems models (PlaSMo) [92].

3.5. From visualization to modelling

Our analysis here was model-assisted but manual, so data visualization was important. For example, phase-plane diagrams can reveal conditional pairwise interactions including subtle effects at low RNA levels, such as the correlation of PRR9 and ELF4 expression in the lhy cca1 double mutant under LL (figure 6d–f). In contrast, PRR9 and ELF4 expression are uncorrelated or anticorrelated in the WT under LD. Such changes in dynamics are important in forming hypotheses during model development. Expert modelling has a subjective element. Objective machine-learning methods can also to contribute to hypothesis generation [93], though understanding such a conditionally connected network (electronic supplementary material, figure S2) is challenging by any approach [94].

Dense transcriptional regulatory interactions might be general for plant environmental response pathways [95], justifying investment in infrastructure to support their analysis. Mathematical models can powerfully express hypotheses about such circuits, so long as the starting model adequately recapitulates most data. Qualitatively, the variation among our datasets was smaller than the departure of the model simulations from the data (figures 3 and 9). The existing circadian clock models are therefore equally applicable to the several growth conditions tested, at least in leaf tissue.

The transition from LD to LL is one case where the model departed from the data, to which it had not previously been constrained (also noted in references [54,88]). The P2011.2.1 model’s 2 h late phase in LL (figure 9) is caused by the slower degradation of PRR proteins in the light than in the dark [16,45]. Without a dark night to reduce PRR levels, their slow degradation delays the rise in LHY/CCA1 on the first cycle in LL in the model. PRR9, PRR7 and PRR5 RNA levels are reduced in the second cycle in LL in both model and data (figures 5c–e and 9e), restoring an approximately 24 h period in subsequent cycles. It is reassuring but not surprising that re-optimization of the model could better match this behaviour, but the models’ detailed behaviour is non-trivial. Reducing the levels of PRR proteins in the new parameter set advanced the phase of the first peak in LL. Simplified models that included only the PRR protein changes also reduced the effect of the PRRs on the period of the clock in constant light (data not shown), contradicting the data. The re-optimization allowed multiple parameter changes to advance the phase of the P2011.2.1 model under LL while retaining the observed effects of PRRs on clock period, such as the short period of the toci1 mutant (figure 9e).

Most significantly, this result was obtained using tools designed to be accessible to biological researchers with no specialist computing or mathematical skills. Development of P2011.2.1 required no new programming, nor the hand-crafted cost functions that were used to optimize previous models [25,83–85], nor the laborious, expert parameter exploration used to construct its parent models [10,14,96]. Our intention was that the scarcity of these skills should no longer present an insuperable barrier, though of course they remain beneficial, not least to keep abreast of relevant method development [80]. To test whether this approach could assist new model development, as well as adjustment of an existing model, we repeated the parameter search within a wide range of values and/or after setting P2011.2.1 model parameters to nominal values. Greater computational power is required when there are fewer constraints on the model’s parameter values; however, viable solutions were identified (table 2) and suitable computing resources are increasingly accessible [97]. The approach and infrastructure presented here allow a wider range of biologists to engage with complicated models, which will be essential tools to understand the mechanisms and physiological functions of complex biological networks.


Competing interests. We declare we have no competing interests.

Funding. Supported by awards from UK BBSRC and EPSRC (ROBuST BB/F005237/1 and SynthSys BB/D019621/1) and from the European Commission (FP7 collaborative project TiMet, contract 245143). This work made use of the facilities of HECToR, the UK’s national high-performance computing service, which was provided by UoE HPCs Ltd at the University of Edinburgh, Cray Inc. and NAG Ltd, and funded by the Office of Science and Technology through EPSRC’s High End Computing Programme.

Acknowledgements. We are grateful to Gavin Steel and Kelly Stewart for expert technical assistance, Martin Beaton and Richard Adams for support of BioDare and SBSI, Uziel Urquiza for preparing SBML files, David Rand for insightful discussion, and members of the A.J.M. and K.J.H. laboratories, who curated data for BioDare.

Appendix A

A.1. Experimental procedures

Experimental methods were similar or identical to published protocols [33,88,98], as detailed in the electronic supplementary material. Statistical significance of comparisons reported in the results is from two-tailed t-tests compared with the ciliate WT plants at the same timepoint, unless otherwise stated. Homoscedasticity is assumed, because all comparisons reported are within individual datasets for the same PCR primers. Significance is not corrected for multiple comparisons (reducing significance), nor for support from neighbouring timepoints or replication across cycles or studies (which can increase significance).
A.2. BioDare and computational methods

The BioDare online resource (www.biodare.ed.ac.uk) uses a desktop application to prepare metadata (describe experiments). The XML editor PEDRO [99] was customized for each experimental protocol to speed up metadata entry, as each experiment can comprise several hundred samples. Numerical data are uploaded in a spreadsheet-compatible format, with the XML metadata, and stored in a relational database. Password-protected access allows controlled data sharing or public dissemination. Searching the metadata (by genotype, marker, etc.) allows aggregation of similar data from multiple sources, followed by secondary processing (detrending, normalization, averaging), visualization (figure 7a) and download. Rhythm analysis in BioDare was recently described [52,53]. Model optimization used SBSVISUAL v. 1.4.5 [29] and SBSINUMERICS v. 1.2 (see electronic supplementary material, Methods). Graphical network diagrams used SBGN-ED in VANTED [38].

A.3. Data, network diagram, model and code accessibility


A.3.1. RNA expression profile data

The RNA datasets reported here are publicly available from BioDare with the permanent data identifiers listed below, using login name ‘public’ with password ‘public’. Numbers below match figure 1.


BioDare accessions:

- ROBuST sd for CCA1, BioDare accession 12820611467827, https://www.biodare.ed.ac.uk/robust/ShowExperiment.action?experimentId=12820611467827
- ROBuST sd for LHY, BioDare accession 3492, https://www.biodare.ed.ac.uk/robust/ShowExperiment.action?experimentId=3492
- ROBuST sd for PRR9, BioDare accession 12820610743262, https://www.biodare.ed.ac.uk/robust/ShowExperiment.action?experimentId=12820610743262
- ROBuST sd for PRR7, BioDare accession 12820611319996, https://www.biodare.ed.ac.uk/robust/ShowExperiment.action?experimentId=12820611319996
- ROBuST sd for PRR5, BioDare accession 12820611188065, https://www.biodare.ed.ac.uk/robust/ShowExperiment.action?experimentId=12820611188065
- ROBuST sd for TOC1, BioDare accession 12820611587928, https://www.biodare.ed.ac.uk/robust/ShowExperiment.action?experimentId=12820611587928
- ROBuST sd for GI, BioDare accession 12820606741450, https://www.biodare.ed.ac.uk/robust/ShowExperiment.action?experimentId=12820606741450

- ROBuST sd for LUX, BioDare accession 12820610913763, https://www.biodare.ed.ac.uk/robust/ShowExperiment.action?experimentId=12820610913763
- ROBuST sd for CAB2, BioDare accession 13228354371807, https://www.biodare.ed.ac.uk/robust/ShowExperiment.action?experimentId=13228354371807
- ROBuST sd for ELF4, BioDare accession 12962296599986, https://www.biodare.ed.ac.uk/robust/ShowExperiment.action?experimentId=12962296599986
- ROBuST sd for ELF3, BioDare accession 12962294335805, https://www.biodare.ed.ac.uk/robust/ShowExperiment.action?experimentId=12962294335805


(7) M.M. Southern and A.J. Millar (2015). Clock RNA timeseries in wild-type and mutant plants under red light. Independent biological replicates are presented for this experiment, with the same genotypes and markers tested in replicates (1 + 2) and (3 + 4):

- Southern sd replicate 1, BioDare accession 1322829828040, https://www.biodare.ed.ac.uk/robust/ShowExperiment.action?experimentId=1322829828040
- Southern sd replicate 2, BioDare accession 13227619871305, https://www.biodare.ed.ac.uk/robust/ShowExperiment.action?experimentId=13227619871305
- Southern sd replicate 3, BioDare accession 1322835705348, https://www.biodare.ed.ac.uk/robust/ShowExperiment.action?experimentId=1322835705348
- Southern sd replicate 4, BioDare accession 13228357183121, https://www.biodare.ed.ac.uk/robust/ShowExperiment.action?experimentId=13228357183121

A.3.2. Code

BioDare and SBSI are open-source and available from Sourceforge (www.sourceforge.net).

BioDare: Sourceforge project, http://sourceforge.net/projects/biodare/. The online resource is available at www.biodare.ed.ac.uk.
A.3.3. Graphical network diagram


A.3.4. Models


The Arabidopsis clock models below are available from the PlaSMo repository (www.plasmo.ed.ac.uk) and will also be submitted to Biomeodels when the present publication has a digital identifier. The model versioning convention is described in the electronic supplementary material.


References


28. Bastow R, Leenelli S. 2010 Sustainable digital infrastructure. Although databases and other online resources have become a central tool for biological research, their long-term support and maintenance is far from secure. EMBO Rep. 11, 730–734. (doi:10.1038/embor.2010.145)


Supporting Information

Contents
Supporting Information ........................................................................................................ 1
Supplementary Methods ....................................................................................................... 2
  Plant materials .................................................................................................................... 2
  Growth, harvesting, RNA extraction and analysis ............................................................. 2
  Evaluation of datasets ....................................................................................................... 5
Supplementary Analysis ....................................................................................................... 6
  Interpretation of Phase-Plane Diagrams .......................................................................... 6
  Model versioning and optimisation .................................................................................. 7
Supplementary Figure Legends ........................................................................................... 8
  Supplementary Figure 1. Overview of the clock gene circuit ........................................ 8
  Supplementary Figure 2. Detailed schema of the P2011 model .................................... 9
  Supplementary Figure 3. Logarithmic plots of clock gene expression in wild-type plants under LD cycles .......................................................... 9
  Supplementary Figure 4. RNA profiles under DD ........................................................... 9
  Supplementary Figure 5. Mutant effects are consistent among data sets ....................... 9
  Supplementary Figure 6. Phase plane diagrams reveal altered regulation in prr7 prr9 mutants .......................................................... 10
  Supplementary Figure 7. Regulation of clock-related genes in low-sugar conditions.... 10
Supplementary Table Legends ........................................................................................... 10
  Supplementary Table 1. mRNA calibration standards ..................................................... 10
  Supplementary Table 2. qRT-PCR primer sequences ..................................................... 11
Supplementary References ................................................................................................. 12
Supplementary Tables ........................................................................................................ 13
  Supplementary Table 1 ..................................................................................................... 13
  Supplementary Table 2. qRT-PCR primer sequences ..................................................... 14
Supplementary Figures ...................................................................................................... 17
Supplementary Methods

Plant materials

The biological materials and allele numbers used in each experiment are shown in Figure 1C.

Growth, harvesting, RNA extraction and analysis

ROBuST dataset (Experiment 1 of Figure 1C)

Seedlings were grown on half-strength Murashige & Skoog media (MS) with 1.2% agar, without added sucrose. Following 7 days of entrainment under LD cycles of 12h L:12h D at 22°C, seedlings were transferred to 17°C on the same light regime, and on the fifth day at 17°C seedlings were harvested every two hours into 1 ml of RNAlater (Ambion; Austin, Texas, USA). Lighting was 60 μmol/m²/s white fluorescent light provided in a Binder (Tuttlingen, Germany) temperature-controlled incubator. Biological triplicates were independently sampled. Samples at ZT0 were taken just before lights-on, and samples at ZT12 just after lights-off. All sampling during lights-off was conduction in a temperature-controlled dark room. Total RNA was extracted (GE Healthcare Illustra RNAspin 96 RNA Isolation Kit, Buckinghamshire, UK) according to the manufacturer’s instructions. cDNA was synthesized from 1 μg of total RNA (Invitrogen SuperScript VILO cDNA Synthesis kit, Paisley, Scotland) and diluted 1:10 in H₂O. qRT-PCR reactions were carried out and analysed as previously described (1) in technical triplicate. Expression values were normalized to the control transcript ACTIN 7 (ACT7), amplified with primers 5’ CAGTGTCTGGATCGGAGGAT 3’ and 5’ TGAACAATCGATGGACCTGA 3’.

TiMet datasets

a. Plants grown on soil: TiMet rosette and seedling 1 datasets (Experiments 2 and 3 of Figure 1C)

Arabidopsis thaliana seeds were sown on wet soil, covered with transparent lids and transferred to growth chambers. Plants were grown in 12 h light / 12 h dark cycles with a light intensity adjusted to 160 μmol m⁻² s⁻¹. Temperature was 20°C during the light phase and 18°C during the dark phase. Pots were randomized to decrease positional effect. After a week lids were removed and the excess plants were thinned. After another seven days plants were treated with Nematodes as a biological pest control.

On the 21st day (rosette plants) or 13th day (seedlings) after sowing two biological replicates per time point were harvested and immediately frozen in liquid nitrogen. Each replicate consisted of a pool of 5-7 plants. Sampling was performed with 2h intervals, starting at ZT0, within 5-10 minutes before the given time point. Plant material was ground using a Ball-Mill (Retch, Germany). Around 50 mg of
material from each sample was aliquoted into 2ml microcentrifuge tubes (Eppendorf, Germany). RNA was extracted using the RNeasy Plant Mini Kit (QIAGEN) by following the manufacturer’s instructions. Briefly, the RLC buffer (500µL) was added to the frozen plant powder and the mixture was homogenised. RNA standards were added (see below). RNA was eluted from the RNeasy spin column twice, first with 50 µL and then with 30 µL of RNase-free water. The concentration of RNA was determined using the Nano-Drop ND-1000 UV-Vis spectrophotometer (Nano-Drop Technologies).

b. Plants grown in sterile media: TiMet seedling 2 dataset (Experiment 4 of Figure1C)

Surface-sterilized seeds were sown on half-strength Murashige-Skoog media (MS) with 1.2 % agar without addition of sucrose. After 4 days of stratification in the dark at 4°C plates were transferred to growth chambers and grown for 10 days in 12 light / 12 dark cycles, with light intensity of approx. 100 µmol m⁻²s⁻¹ and temperature of 22°C during the day and 18°C during the night. Seedlings were harvested on the 10th day into 2ml Eppendorf tubes and immediately frozen in liquid nitrogen. For each time point, three independent biological replicates were collected; each included 10-20 seedlings. Samples were taken within 10 – 15 min before the given time point. Frozen seedlings were ground using a Ball-Mill (Retch, Germany). RNA was extracted using the RNeasy Plant Mini Kit (QIAGEN) but using half the volume of all buffers. RNA standards were added (see below). In the last step RNA was eluted from the RNeasy spin column once with 20 µL of RNase-free water. The concentration of RNA was determined using the Nano-Drop ND-1000 UV-Vis spectrophotometer (Nano-Drop Technologies).

c. cDNA preparation and analysis

Contaminating DNA was removed from samples using TURBO DNA-free™ kit (Applied Biosystems) by following the manufacturer’s instructions. The concentration of RNA was determined using Nano-Drop spectrophotometer. 1µg of DNase treated RNA was used for cDNA synthesis. Reverse transcription used the SuperScript III First-Standard Synthesis System Kit (Invitrogen). Briefly, 1µg of DNase treated RNA was mixed with 2µL of oligo(dT) primers (50µM), 2µL of random hexamers (50ng/µL) and 1µL of dNTP mix (10mM). Mixture of a final volume of 11µL was incubated for 5 min at 65°C in a Verti 96 Well Thermal Cycler (Applied Biosystems, Deutschland). After cooling down the mixture for 5 min on ice the extension reaction was performed by adding 0.5µL of reverse transcriptase (SuperScript III RT 200 U/µl), 2µL of 10X First strand synthesis buffer, 4µL of MgCl₂ (25mM), 2µL of DTT (0.1M) and 0.5µL of RNase OUT (40 U/µL). Incubation for 10 min at 25°C was followed by 50 min incubation at 50°C. Next, enzyme was inactivated by heating for 5 min at 85°C. To digest all RNA that could potentially be left in samples 1 µL RNase H (2 U/µL) was added to the mixture and incubated at 37°C for 20 min and then for 5 min at 85°C. Samples were diluted in RNase-free water (dilution 1/5).
The cDNA was used for qRT-PCR reactions to monitor the expression level of 10 core clock genes (LHY, CCA1, TOC1, PRR9, PRR7, PRR5, GI, ELF3, ELF4 and LUX). To control the quality of performed reactions, four pairs of primers encoding housekeeping genes (ACT2 155, ACT2 633, GAPDH 3’, GAPDH 5’) were included to the measurements. Primer sequences are listed in Supplementary Table 2. To allow absolute quantification, 8 pairs of primers from the ArrayControl RNA Spikes (Applied Biosystems) were included (see below). The 22 pairs of primers were dispensed (200nM of each primer) into 384-well plates for using PerkinElmer Evolution P3 Precision Pipetting Platform (PerkinElmer Life Science, Rodgau-Jügesheim, Germany). PCR mix was prepared by adding 0.5µL of cDNA and 2.5µL of Power SYBR Green PCR Master Mix (Applied Biosystems, Deutschland) to already prepared primers. The qRT-PCR reactions were carried out following the recommended thermal profile: 2 min at 50˚C, 10 min at 95˚C, followed by 40 cycles of 95˚C for 15 s and 60 ˚C for 1 min, in an ABI PRISM 7900 HT sequence detection system (Applied Biosystems Deutschland, Darmstadt, Germany). The specificity of the amplifications was tested by heating from 60˚C to 95˚C with a speed of 1.9 ˚C min⁻¹, resulting in melting curves. Data analysis was performed using SDS 2.4 software (Applied Biosystems Deutschland). Data were normalised on the basis of results obtained from measurements of standard material that was included in all performed experiments. The standard samples were collected from Arabidopsis thaliana plants, ecotype Col-0, grown on soil in a glasshouse.

d. Data calibration to absolute values

The transcript data are expressed in RNA copy number per cell. This absolute quantification of the transcripts level was achieved using artificial poly(A)⁺ RNAs (ArrayControl RNA Spikes - Applied Biosystems) as internal standards. Eight different artificial mRNA spikes with known concentration (Supplementary Table 1) were mixed and the same amount of the spikes mix was added to all tested samples at the first step of the RNA extraction procedure (after adding the RLC buffer to the frozen plant material, see above). The efficiency and the Cₜ values of all 8 spikes were used to generate a standard curve for each sample. Each measured data point was plotted against the standard curve to calculate the concentration of mRNA in units of copy number per gram fresh weight (copy number / g FW). The procedure is very similar to our earlier report (2), which also addressed the accuracy of the measurements. Our results are in line with their conclusion that, for a single timepoint with achievable replication, “Precise quantification of changes of [less than] two-fold is difficult using qRT–PCR.”. Adjacent timepoints and successive daily cycles in circadian data provide greater support than the single-timepoint assays in (2). Rigorous assessment of the uncertainties will require more sophisticated statistical models, as the system’s behaviour is not independent at successive, 2-h timepoints.

To determine the cell number per g FW, the genome copy number per gFW was quantified. For this purpose artificial DNA spikes were added to the extract during DNA extraction (at the beginning of the protocol). Next, PCR reactions with primers for genes that are known to have only a single copy
per genome (e.g. mitochondrial 26S ribosomal RNA protein, NADH dehydrogenase subunit 9, NAD(P)H dehydrogenase subunit H protein) were performed. The ploidy level of the samples was estimated using flow cytometry. Finally, the genome copy number per g FW was divided by the ploidy level. The resulting number (25,000,000 copy number/g FW) was used to express metabolite and transcript data.

**McWatters dataset (Experiment 5 of Figure 1C)**

Seedlings were grown on sterile media containing 3% sucrose. Entire seedlings including roots were harvested every 3h. RNA was extracted, amplified and analysed, all using methods previously described (3).

**Edwards dataset (Experiment 6 of Figure 1C)**

Seedlings were grown on sterile media containing 3% sucrose. Entire seedlings including roots were harvested every 2h. RNA was extracted and analysed as described (4).

**Southern dataset (Experiment 7 of Figure 1C)**

Seedlings were grown on sterile media containing 3% sucrose, harvested every 3h with additional sampling times around light-dark transitions; RNA was extracted, amplified and analysed, as described (5).

**Evaluation of datasets**

The larger data sets reported here were acquired using automated, quantitative RT-PCR platforms that were established as shared facilities for research centres in Golm and Edinburgh. The experiments remained laborious and costly, despite the automation and reduced reagent costs from small assay volumes; further miniaturisation of these assays shows promise for future studies (as in 6). Large-scale data cannot usually be acquired with the same uniformity as small-scale datasets. Replacing individual measurements with supplementary experiments to improve technical quality is infeasible, because the broad scope of genotypes, conditions and timepoints in the original studies makes such experiments expensive relative to the small number of improved data points, and introduces a further source of variation. Thus we note technical variation in order to inform subsequent analysis, reinforcing the point that only a small proportion of the data were affected. RNAs with low amplitude regulation (TOC1, ELF3) or low expression levels (PRR9) are most obviously affected. Automated liquid handling avoided the sample-to-sample variation of manual pipetting in the earlier, Southern data (Supplementary Figure 5E, 5F) but risked introducing larger-scale effects. Three issues affected multiple samples. One timepoint (ZT24) in the ROBuST data was discarded for several genotypes due
to a harvesting error; ZT24 data are expected to be very similar to ZT0 but the tissue is one day older. Some technical replicates in ROBuST were discarded and repeated wholesale when a trend was discovered across multiple samples, due to a progressive liquid handling error in the robot used in PCR setup. The first Col-0 samples in the TiMet rosette dataset showed high biological variability due to harvesting or calibration, so the entire first 24h of the timeseries were replaced by a supplementary experiment (shown as experiment 2B in Figure 1C).

No molecular phenotype was detected when the *toc1* mutant was tested under red light in the Southern data set (data not shown). The strong *toc1-2* mis-sense allele that was used in this study causes a splicing defect (7). The expected change in the size of the *TOC1* transcript was detected using the method described (7) on our RNA samples, confirming their genotype (8). *toc1-2* seedlings do show hypocotyl elongation defects when grown under red light on media without exogenous sucrose (9), so the absence of molecular phenotype is possibly due to 3% sucrose in our media. The *toc1-9* and *toc1-100* alleles tested under white light in the ROBuST and TiMet data sets gave the expected molecular phenotypes.

We found no consistent effect on the RNA profiles from normalisation to endogenous control RNAs. TiMet rosette data were normalised to absolute, external standards (Figure 1) or to internal controls (Figure 3), for comparison to the profiles from other data sets. The profiles were very similar in both analyses, even for features that varied across experiments. In the GI profile, for example, the minor peak at ZT2 was 20% of the major peak at ZT8 in Figure 1 in Col in Figure 1 (18% in Figure 3), or 11% (10%) in Ws. The 1-2% differences were much less than the biological variation (see Figure 1).

**Supplementary Analysis**

**Interpretation of Phase-Plane Diagrams**

Phase plane diagrams plot the levels of two components against each other (Figure 7), exactly as in a scatter plot used to test for correlation. The data in scatter plots do not usually have a meaningful sequence. Timeseries do, so the data points in a phase plane diagram are connected by a line that represents their time sequence. Data symbols for ZT0 and ZT12 are enlarged to provide time references, because sample times are not directly plotted. Rhythmic profiles in phase lead to a diagonal plot with a positive gradient, for example for GI and TOC1 expression (Figure 7A), whereas rhythms 90 or 270 degrees out of phase produce a circular plot. GI and TOC1 expression reached peak and trough levels at overlapping timepoints, elongating the plot along the positive diagonal. The plots form an open shape rather than a line, because GI rose without (before) TOC1, especially in Col plants of the TiMet and ROBuST data sets, without exogenous sucrose. High TOC1 occurred after
high GI, particularly in Ws plants of the TiMet data sets. This relative timing is of course easy to see in conventional timeseries plots, and has been noted in many past studies.

Rhythms 180 degrees out of phase yield a right-angled plot on a linear scale, or a diagonal of gradient -1 on a logarithmic scale, illustrated in the plots of \textit{CCA1} and \textit{TOC1} RNA (Figures 7B, 7C). The logarithmic scale shows that \textit{CCA1} expression rises over 100-fold along roughly horizontal lines, indicating little effect on \textit{TOC1}. In contrast, the rise in \textit{TOC1} expression directly correlates with falling \textit{CCA1}, resulting in diagonals with negative gradients (Figure 7C). Linear scale plots might best represent potential protein synthesis, which is usually assumed to increase linearly with mRNA level; hence linear plots suggest likely effects on downstream targets. On the other hand, logarithmic scale plots reveal the effects of upstream regulators, even at low RNA levels that are indistinguishable on a linear scale.

The dependence of an interaction of two genes upon a third regulator was strikingly illustrated by plotting the profiles of \textit{PRR9} and \textit{ELF4}, which peak far out of phase in the wild type (Figure 7D). Starting from ZT0, acute light activation of \textit{PRR9} drives peak expression at ZT2-6 while \textit{ELF4} expression is minimal. \textit{ELF4} expression rises to peak at ZT10 as \textit{PRR9} falls, consistent with some contribution to repressing \textit{PRR9}. \textit{ELF4} expression falls at night (red dashed line) while \textit{PRR9} expression is minimal. Data from LL (filled symbols) suggest a negative correlation in the subjective night, when \textit{ELF4} falls as \textit{PRR9} rises. In the \textit{lh y cca1} double mutant under LD cycles (Figure 7E, open symbols), \textit{PRR9} is partially de-repressed at ZT0, but remains out of phase with \textit{ELF4}. However, the two genes are co-regulated for part of the cycle under LL, creating a diagonal with a positive gradient (red dashed line, Figure 6F).

Phase plane diagrams for the \textit{prr7 prr9} double mutant (Figure S6) showed that \textit{ELF4} expression rose while \textit{CCA1} (and \textit{LHY}, not shown) were still significantly expressed at the end of the day, suggesting that not only \textit{CCA1} and \textit{LHY} but also the \textit{PRRs} repress \textit{ELF4} in the wild type. Removing \textit{PRR9} and \textit{PRR7} also revealed simpler correlations of \textit{PRR5} expression with \textit{CCA1} and \textit{ELF4} (Figures S6E, S6H).

\textbf{Model versioning and optimisation}

We have adapted the Community Earth System Model version-naming convention (10) (www2.cesm.ucar.edu/models) to describe version of our clock models. Models are named as Axxxx.Y.Z., where \textit{A} is a letter of the model author’s surname, xxxx is the year of submission, \textit{Y} is a version number marking a significant change in the model that alters its dynamics such as a new
parameter set, Z is a minor change that does not alter the model’s dynamics such as the refactoring of the light input function. The version identifier should link to a unique identifier in at least one model repository.

Model optimisation used SBSIvisual version 1.4.5 (11) and SBSInumerics version 1.2, which are both available from Sourceforge. Model P2011.1.2 (PLM_64, version 4) was optimised using SBSI’s Parallel Genetic Algorithm, with a range of parameter constraints (Table 2). Starting parameter sets were chosen as normal (P2011.2.1) or logarithmic values (P2011.3-P2011.6) within the stated ranges, either randomly (P2011.3-P2011.6) or from a quasi-random Sobol series (P2011.2.1). The departure of each model’s simulations of wild-type and mutants from the normalised TiMet rosette data was evaluated using SBSI’s chi-square cost function, and their departure from measured period values under constant light was evaluated using SBSI’s FFT cost function. SBSInumerics v1.2 estimates the period of simulated timeseries with greater precision than v1.1. SBSInumerics does not specifically cost the amplitude of the simulated timeseries, which are constrained only by the match to the RNA data. The results (Figure 9) suggest that small discrepancies in simulated amplitude are better tolerated in the costing than discrepancies in timing, matching our priorities. Moreover, our choice of Hill coefficients is very conservative (all set at 2), restricting the nonlinearity of simulated gene regulation and tending to reduce the amplitudes of the simulated RNA profiles. The >100-fold relative amplitudes observed for some clock RNAs suggest much greater non-linearity of regulation in vivo.

Optimisation job .599 tested 5000 randomly-generated parameter sets and optimised the best solutions in a population of 200 models over 599 generations; computation time was 6 hours on a server with eight Intel Xeon processing cores. Job t40 tested randomly-generated parameter sets before optimising the best solutions as above; computation time was 8.5 hours using 4096 processing cores on the UK national supercomputing resource HECToR. The parameter constraints and optimised parameter values are listed in the SBML files (see Data, model and code accessibility).

**Supplementary Figure Legends**

**Supplementary Figure 1. Overview of the clock gene circuit**

The clock gene circuit summarised in the Activity-Flow language of SBGN v1.0 (12), with the principal light inputs and regulatory interactions in the P2012 model (13). The repressilator is denoted by green lines; morning loop components are filled yellow; evening loop components are filled blue.
Supplementary Figure 2. Detailed schema of the P2011 model

The clock gene network in the P2011 model (14), represented in the SBGN Process Description language. All connections in the model are depicted, along with the logic inherent in the form of the equations. Coloured edges correspond to different parts of the model: Yellow edges - Light inputs. Green edges - Morning loop. Blue edges - Evening loop. Purple edges – connections among loops.

Supplementary Figure 3. Logarithmic plots of clock gene expression in wild-type plants under LD cycles

Transcript levels in Col-0 and Ws-2 wild types under LD 12:12 were measured by qRT-PCR, in experiment 2 (TiMet ros) including eight external RNA standards to allow absolute quantification in Col-0 and Ws-2 (A, C, E) and in experiment 1 (ROBuST) normalised to the ACTIN7 control in Col-4 and Ws-2 (B, D, F). Data represent transcripts of (A, B) LHY and CCA1, (C, D) PRR9, (E, F) TOC1 and GI. Error bars show SD, for 2-3 biological replicates. Figure 2 shows the same data on a linear scale.

Supplementary Figure 4. RNA profiles under DD

RNA profiles of (A) CCA1 and (B) GI expression from the Edwards dataset for seedlings on sucrose (4), under LD cycles of 6, 12 or 18h photoperiod followed by DD (legend, panel B). RNA profiles from the TiMet rosette dataset for plants in 12h photoperiod, transferred to DD, for (C) PRR7, (D) PRR9 expression. Higher trough levels are observed in DD than under LD, often more than ten-fold higher. (A), (B) Data shown are relative to the ACTIN2 control; (C)-(D) data are calibrated to RNA copy number per cell. Means of two biological replicates per timepoint are shown, error bar is SD. Open box, light interval; filled box, dark interval, A and B, coloured as for legend.

Supplementary Figure 5. Mutant effects are consistent among data sets.

RNA profiles from the ROBuST dataset for seedlings under LD cycles show (A) the slower fall of PRR5 expression in the gi-11 mutant compared to wild type, persisting from ZT10-16h (logarithmic scale, B); (C) earlier and lower peak expression of TOC1 in the lhy cca1 double mutant; (D) lower expression of ELF3 in the gi mutant, especially at ZT8-10h; (E) persistent expression of CCA1 in the prr7 prr9 double mutant, with an early rise in prr7 and delayed fall in prr9. RNA profiles from the Southern data for elf3 and elf4 mutant seedlings under red light-dark cycles show (F) persistently high and noisy expression of GI; (G) expression of CCA1 that falls from the level in Ws
to elf4, which remains rhythmic in constant conditions, and falls further to elf3, which is arrhythmic.

A-D show means of three biological replicates per timepoint, error bars are SEM. E, F show one of two replicate experiments with similar results. Plots in panels C and D are from BioDare.

**Supplementary Figure 6. Phase plane diagrams reveal altered regulation in prr7 prr9 mutants.**

RNA profiles of Figure 5 are represented as phase plane diagrams on logarithmic scales, plotting data for ELF4 and CCA1 (A) in wild-type Col plants under LD and LL and (B) in Col plants under LD and prr7 prr9 double mutants under LD and LL, with (C) the LD data for both genotypes on linear scales. (D)-(E) show data for PRR5 and CCA1 (D) in wild-type Col plants under LD and LL and (E) in Col plants under LD and prr7 prr9 double mutants under LD and LL. (E) Red dashed line marks anticorrelated levels during the subjective night in the double mutant in LL.

(F)-(H) show data for ELF4 and PRR5 (F) in wild-type Col plants under LD and LL, (G) in Col plants under LD and prr7 prr9 double mutants under LD and (H) for prr7 prr9 double mutants under LD and LL. (H) Red dashed lines mark highly correlated rise and fall of PRR5 and ELF4 levels in the double mutant under LL, whereas the relationship was more complex in the wild type. Larger markers indicate ZT0(24) and ZT12(36) datapoints in LD (LL), arrows indicate the direction of time.

**Supplementary Figure 7. Regulation of clock-related genes in low-sugar conditions.**

RNA microarray data (15, 16) displayed by the online tool (http://mapman.mpimp-golm.mpg.de/supplement/xn/) from treatments with light, CO2-free air (ΔCO2), DD (eXtended Night), or the starchless pgm mutant, for (A) LHY, PRR7 (green), PRR5 and TOC1; (B) ZTL, FKF1 and LKP2. (A) Arrows mark higher PRR7 levels in sugar-limiting DD and pgm relative to control in LD cycle, repression by re-supply of high exogenous sugar (3 Suc) but less effect from resupply of normal air (350ppm CO2). (B) Arrows mark higher levels of ZTL, FKF1 and LKP2 RNA in DD and pgm.

**Supplementary Table Legends**

**Supplementary Table 1. mRNA calibration standards.**

Concentration of artificial mRNA spikes (number of copies per extract) added to each sample as an internal standard to allow the quantification of RNA copy number.
Supplementary Table 2. qRT-PCR primer sequences.

The PCR primer sequences used in the TiMet studies are listed below. The PCR primer sequences used in the ROBuST data set are listed in the public record for each transcript on BioDare (please see Data Accessibility), and the primer sequences for other studies are listed in their original publications and in the BioDare records (please see Data Accessibility).
Supplementary References


**Supplementary Tables**

**Supplementary Table 1**

Concentration of artificial mRNA spikes (number of copies per extract) added to each sample as an internal standard to allow the quantification of RNA copy number.

<table>
<thead>
<tr>
<th></th>
<th>Spike 1</th>
<th>Spike 2</th>
<th>Spike 3</th>
<th>Spike 4</th>
<th>Spike 5</th>
<th>Spike 6</th>
<th>Spike 7</th>
<th>Spike 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Copies per extract</td>
<td>$6.08 \times 10^9$</td>
<td>$1.52 \times 10^9$</td>
<td>$4.56 \times 10^8$</td>
<td>$1.14 \times 10^8$</td>
<td>$2.94 \times 10^7$</td>
<td>$3.04 \times 10^5$</td>
<td>$2.58 \times 10^6$</td>
<td>$7.60 \times 10^5$</td>
</tr>
</tbody>
</table>
**Supplementary Table 2. qRT-PCR primer sequences.**

The PCR primer sequences used in the TiMet studies are listed below, followed by the PCR primers for the Southern dataset. The PCR primer sequences used in the ROBuST data set are listed in the public record for each transcript on BioDare (please see Data Accessibility), and the primer sequences for other studies are listed in their original publications and/or in the BioDare records.

Primers for the TiMet datasets.

<table>
<thead>
<tr>
<th>Product</th>
<th>CODE</th>
<th>SEQUENCE</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spike 1</td>
<td>H251-A1780s1-f1</td>
<td>GATGCCGACCACATCATTTAG</td>
<td></td>
</tr>
<tr>
<td></td>
<td>H252-A1780s1-r1</td>
<td>GTGAGGTAATGTCGCTTC</td>
<td></td>
</tr>
<tr>
<td>Spike 2</td>
<td>H253-A1780s2-f1</td>
<td>ATTGCCTGCCATATAAC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>H254-A1780s2-r1</td>
<td>GCTGGGATCAAGGAGGAAG</td>
<td></td>
</tr>
<tr>
<td>Spike 3</td>
<td>H255-A1780s3-f1</td>
<td>GATCGTTTGCCCTGCATTACC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>H256-A1780s3-r1</td>
<td>GAGAGCGTCAAGCATAACC</td>
<td></td>
</tr>
<tr>
<td>Spike 4</td>
<td>H257-A1780s4-f1</td>
<td>TACGTCGCAACAACCACAATC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>H258-A1780s4-r1</td>
<td>CAGCGCCACATAACCCAACACT</td>
<td></td>
</tr>
<tr>
<td>Spike 5</td>
<td>H302-S5f-2</td>
<td>ACAAAAGGAGCGTTGAAAAC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>H301-S5r-1</td>
<td>TAGTGTCGTACGCCATACCC</td>
<td></td>
</tr>
<tr>
<td>Spike 6</td>
<td>Spike-6f-2</td>
<td>CGCAAAGTCTCTCTCTCTTTGG</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Spike-6r-2</td>
<td>CAGTAGCCATTGCCGGAAGAT</td>
<td></td>
</tr>
<tr>
<td>Spike 7</td>
<td>H313-S7f-3</td>
<td>CTGAAACCAGACTGCACATGG</td>
<td></td>
</tr>
<tr>
<td></td>
<td>H314-S7r-3</td>
<td>ACGCTCTAGGGGTTTGTATAT</td>
<td></td>
</tr>
<tr>
<td>Spike 8</td>
<td>H315-S8f-3</td>
<td>TCAATTAAGCGGAAGGCAAT</td>
<td></td>
</tr>
<tr>
<td></td>
<td>H316-S8r-3</td>
<td>AAATTTCCAGCACCAGTCAG</td>
<td></td>
</tr>
<tr>
<td>GAPDH 3’</td>
<td>H278-GAPDH3</td>
<td>TTGGTGACAAACAGTTCAAGCA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>H279-GAPDH3</td>
<td>AAACCTTCCTGAATGCAAT</td>
<td></td>
</tr>
<tr>
<td>GAPDH 5’</td>
<td>H280-GAPDH5</td>
<td>TCTCGATCTCAATTCGCAAAA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>H281-GAPDH5</td>
<td>CGAACCGTTGATCCCGATTC</td>
<td></td>
</tr>
<tr>
<td>ACT2 155</td>
<td>H282-155-155-ACT2</td>
<td>AACTCTCCCCGCTATGTATGTCGC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>H283-155-155-ACT2</td>
<td>CAATACCCTTTGTAAGCACCAGT</td>
<td></td>
</tr>
<tr>
<td>ACT2 633</td>
<td>H284-633-ACT2</td>
<td>ACTTTACCAGCCTGTTTGA</td>
<td></td>
</tr>
<tr>
<td>Primers</td>
<td>Forward Sequence</td>
<td>Reverse Sequence</td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>----------------------------------------</td>
<td>----------------------------------------</td>
<td></td>
</tr>
<tr>
<td>H285-633-</td>
<td>ACGATTGTTGAATATCATCAG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACT2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LHY</td>
<td>TGCCCTCAAAGCTTTTCGCTTCT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F019-LHY-f</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F020-LHY-r</td>
<td>GTCTGCAGCACAAAAGGATCTCGGCT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCA1</td>
<td>TCCAATGCACGCGCAGTAGA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F027-CCA1-f</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F028-CCA1-r</td>
<td>AGGCAATTCGACCCCTCGTCAGCA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PRR9</td>
<td>AGCTAGCAGAAACAACGTCTCGAGT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F037-PRR9-f</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F038-PRR9-r</td>
<td>CGTCTGAATTCACGGTTCGCACGA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PRR9</td>
<td>GATTGGTGAAATTCGACAAGC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JF309-PRR9-F</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>primers from A. Millar</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JF310-PRR9-R</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PRR7</td>
<td>AGAGGTGCTTCGGAAAGAGGTACGA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F045-PRR7-f</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F046 -PRR7-r</td>
<td>ACGCACAATTGGCCTCGCACT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PRR5</td>
<td>TGCAATGGCTCCTGCTTACACTTC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F051-PRR5-f</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F052-PRR5-r</td>
<td>ACTGTACTCATGCGGGCTAACGGA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOC1</td>
<td>TGATGGATCGGTTCTCTTCGACCA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F073-TOC1-f</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F074-TOC1-r</td>
<td>TGAGGCATCATGGCTGCTGATTGC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LUX</td>
<td>CAGCGGTAATGTTGGAGTGCCGAT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F077-LUX-f</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F078-LUX-r</td>
<td>TGGCATCTGACATCTCTGGGTCT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GI</td>
<td>TGGCGGCAACTGATGGAATGCT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F083-GI-f</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F084-GI-r</td>
<td>TGCTTTGGCCGTGGCTTCAAAGT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ELF4</td>
<td>AGTTTCTCGTGCCGCTTTTCACGGT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F091-ELF4-f</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F092-ELF4-r</td>
<td>TAAAGCTCTAGTCCGCGACGACCA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ELF3</td>
<td>ACAACAAAGAGATGGGGGAGGAGGTGAC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F099-ELF3-f</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F100-ELF3-r</td>
<td>ACTCGCGAGCTTTGGCGGTGTA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ELF3</td>
<td>CATTCGACAGCGTTGAGTGAGG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JF576-ELF3-F</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>from Coluccio et al. J. Exp.Bot, 2011</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JF577-ELF3-R</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Primers for the Southern dataset (8).

<table>
<thead>
<tr>
<th>Product</th>
<th>Forward primer</th>
<th>Reverse primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACT2</td>
<td>CAGTGTCCTGGATGGAGGAT</td>
<td>TGAACATCGATGGAGACTGA</td>
</tr>
<tr>
<td>LHY</td>
<td>CAACAGCAGCAACATGCAACTAC</td>
<td>AGAGAGCTGAAACGCTATACGA</td>
</tr>
<tr>
<td>CCA1</td>
<td>CTGTGTCTGACGGGTCGAA</td>
<td>ATATGTAACCTTTGGCAGAATACCT</td>
</tr>
<tr>
<td>GI</td>
<td>AATTCAGACCGGAGCTATT</td>
<td>GGTGCTCTGCTGCAAGGACTT</td>
</tr>
<tr>
<td>TOC1</td>
<td>ATCTTCGAGAATCCCTGAGTA</td>
<td>GCACCTAGTTCAGCAACTTTACA</td>
</tr>
<tr>
<td>ELF3</td>
<td>GGAAAGCCATTGCTAAACATCA</td>
<td>ATCCCGTGGATGCAAGAATAAGT</td>
</tr>
<tr>
<td>ELF4</td>
<td>CGCAATCACAATCGAGAATG</td>
<td>AATGTTCCGTTGAGTTCTGAAATC</td>
</tr>
</tbody>
</table>
Supplementary Figure 1. Overview of the clock gene circuit
The clock gene circuit summarised in the Activity-Flow language of SBGN v1.0 (Le Novere et al., 2009), with the principal light inputs and regulatory interactions in the P2012 model (Pokhilko et al., 2013). The repressilator is denoted by green lines; morning loop components are filled yellow; evening loop components are filled blue.
Supplementary Figure 2. Detailed schema of the P2011 model

The clock gene network in the P2011 model (Pokhilko et al., 2012), represented in the SBGN Process Description language. All connections in the model are depicted, along with the logic inherent in the form of the equations. Coloured edges correspond to different parts of the model: Yellow edges - Light inputs. Green edges - Morning loop. Blue edges - Evening loop. Purple edges – connections among loops.
Supplementary Figure 3

A. Time after dawn (hour) vs. copy no./gfrw

B. Time after dawn (hour) vs. expression arb. units

C. Time after dawn (hour) vs. copy no./gfrw

D. Time after dawn (hour) vs. expression arb. units

E. Time after dawn (hour) vs. copy no./gfrw

F. Time after dawn (hour) vs. expression arb. units

G. Time after dawn (hour) vs. copy no./gfrw

H. Time after dawn (hour) vs. expression arb. units

I. Time after dawn (hour) vs. copy no./gfrw

J. Time after dawn (hour) vs. expression arb. units
**Supplementary Figure 3. Logarithmic plots of clock gene expression in wild-type plants under LD cycles**

Transcript levels in Col-0 and Ws-2 wild types under LD 12:12 were measured by qRT-PCR, in experiment 2 (TiMet ros) including eight external RNA standards to allow absolute quantification in Col-0 and Ws-2 (A, C, E) and in experiment 1 (ROBuST) normalised to the ACTIN7 control in Col-4 and Ws-2 (B, D, F). Data represent transcripts of (A, B) LHY and CCA1, (C, D) PRR9, (E, F) TOC1 and GI. Error bars show SD, for 2-3 biological replicates. Figure 2 shows the same data on a linear scale.
Supplementary Figure 4. RNA profiles under DD
RNA profiles of (A) CCA1 and (B) GI expression from the Edwards dataset for seedlings on sucrose (Edwards et al., 2010) under LD cycles of 6, 12 or 18h photoperiod followed by DD (legend, panel B). RNA profiles from the TiMet rosette dataset for plants in 12h photoperiod, transferred to DD, for (C) PRR7, (D) PRR9 expression. Higher trough levels are observed in DD than under LD, often more than ten-fold higher. (A), (B) Data shown are relative to the ACTIN2 control; (C)-(D) data are calibrated to RNA copy number per cell. Means of two biological replicates per timepoint are shown, error bar is range. Open box, light interval; filled box, dark interval, A and B, coloured as for legend.
Supplementary Figure 5. Mutant effects are consistent among data sets.
RNA profiles from the ROBuST dataset for seedlings under LD cycles show (A) the slower fall of PRR5 expression in the gi-11 mutant compared to wild type, persisting from ZT10-16h (logarithmic scale, B); (C) earlier and lower peak expression of TOC1 in the lhy cca1 double mutant; (D) lower expression of ELF3 in the gi mutant, especially at ZT8-10h; (E) persistent expression of CCA1 in the prr7 prr9 double mutant, with an early rise in prr7 and delayed fall in prr9. RNA profiles from the Southern data for elf3 and elf4 mutant seedlings under red light-dark cycles show (F) persistently high and noisy expression of GI; (G) expression of CCA1 that falls from the level in Ws to elf4, which remains rhythmic in constant conditions, and falls further to elf3, which is arrhythmic. A-D show means of three biological replicates per timepoint, error bars are SEM. E, F show one of two replicate experiments with similar results. Plots in panels C and D are from BioDare.
Supplementary Figure 6. Phase plane diagrams reveal altered regulation in prr7 prr9 mutants.

RNA profiles of Figure 5 are represented as phase plane diagrams on logarithmic scales, plotting data for ELF4 and CCA1 (A) in wild-type Col plants under LD and LL and (B) in Col plants under LD and prr7 prr9 double mutants under LD and LL, with (C) the LD data for both genotypes on linear scales. (D)-(E) show data for PRR5 and CCA1 (D) in wild-type Col plants under LD and LL and (E) in Col plants under LD and prr7 prr9 double mutants under LD and LL. (E) Red dashed line marks anti-correlated levels during the subjective night in the double mutant in LL. (F)-(H) show data for ELF4 and PRR5 (F) in wild-type Col plants under LD and LL, (G) in Col plants under LD and prr7 prr9 double mutants under LD and (H) for prr7 prr9 double mutants under LD and LL. (H) Red dashed lines mark highly correlated rise and fall of PRR5 and ELF4 levels in the double mutant under LL, whereas the relationship was more complex in the wild type. Larger markers indicate ZT0(24) and ZT12(36) datapoints in LD (LL), arrows indicate the direction of time.
Supplementary Figure 7. Regulation of clock-related genes in low-sugar conditions.
RNA microarray data (Blasing et al., 2005; Usadel et al., 2008) displayed by the online tool (http://mapman.mpimp-golm.mpg.de/supplement/xn/) from treatments with light, CO₂-free air (ΔCO₂), DD (eXtended Night), or the starchless pgm mutant, for (A) LHY, PRR7 (green), PRR5 and TOC1; (B) ZTL, FKF1 and LKP2. (A) Arrows mark higher PRR7 levels in sugar-limiting DD and pgm relative to control in LD cycle, repression by re-supply of high exogenous sugar (3 Suc) but less effect from resupply of normal air (350ppm CO₂). (B) Arrows mark higher levels of ZTL, FKF1 and LKP2 RNA in DD and pgm.