THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

Testing the inferred transcription rates of a dynamic, gene network model in absolute units

Citation for published version:
Urquiza-García, U & Millar, AJ 2021, 'Testing the inferred transcription rates of a dynamic, gene network model in absolute units', *In Silico Plants*. https://doi.org/10.1093/insilicoplants/diab022

Digital Object Identifier (DOI):
10.1093/insilicoplants/diab022

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
In Silico Plants

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.

OPEN ACCESS

Download date: 23. Feb. 2022

Testing the inferred transcription rates of a dynamic, gene network model in absolute units

Uriel Urquiza-García[1,2] and Andrew J. Millar*[1]

[1]SynthSys and School of Biological Sciences, C. H. Waddington Building, University of Edinburgh, King's Buildings, Edinburgh EH9 3BF, United Kingdom

[2]Institute of Synthetic Biology, University of Düsseldorf, Düsseldorf, Germany, CEPLAS-Cluster of Excellence on Plant Sciences, Düsseldorf, Germany

ORCID and email:

Uriel Urquiza-García 0000-0002-7975-5013, urquizag@hhu.de

Andrew Millar 0000-0003-1756-3654, andrew.millar@ed.ac.uk

*Corresponding author:

Andrew J. Millar

SynthSys, C.H. Waddington Building, University of Edinburgh, Max Born Crescent, Edinburgh EH9 3BF, Scotland, UK,

telephone: +44 (0)131 651 3325

fax: +44 (0)131 651 9068

email: andrew.millar@ed.ac.uk

© The Author(s) 2021. Published by Oxford University Press on behalf of the Annals of Botany Company.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from https://academic.oup.com/insilicoplants/advance-article/doi/10.1093/insilicoplants/diab022/6342117 by guest on 11 August 2021

Abstract

The circadian clock coordinates plant physiology and development. Mathematical clock models have provided a rigorous framework to understand how the observed rhythms emerge from disparate, molecular processes. However, models of the plant clock have largely been built and tested against RNA timeseries data in arbitrary, relative units. This limits model transferability, refinement from biochemical data and applications in synthetic biology. Here, we incorporate absolute mass units into a detailed model of the clock gene network in *Arabidopsis thaliana*. We re-interpret the established P2011 model, highlighting a transcriptional activator that overlaps the function of REVEILLE 8/LHY-CCA1-LIKE 5. The U2020 model incorporates the repressive regulation of *PRR* genes, a key feature of the most detailed clock model KF2014, without greatly increasing model complexity. We tested the experimental error distributions of qRT-PCR data calibrated for units of RNA transcripts/cell and of circadian period estimates, in order to link the models to data more appropriately. U2019 and U2020 models were constrained using these data types, recreating previously-described circadian behaviours with RNA metabolic processes in absolute units. To test their inferred rates, we estimated a distribution of observed, transcriptome-wide transcription rates (Plant Empirical Transcription Rates, PETR) in units of transcripts/cell/hour. The PETR distribution and the equivalent degradation rates indicated that the models' predicted rates are biologically plausible, with individual exceptions. In addition to updated clock models, FAIR data resources and a software environment in Docker, this validation process represents an advance in biochemical realism for models of plant gene regulation.

keywords: systems biology, circadian rhythms, biological clocks, mathematical modelling, parameter estimation, RNA metabolism, transcriptional regulation, *Arabidopsis thaliana*.

Downloaded from https://academic.oup.com/insilicoplants/advance-article/doi/10.1093/insilicoplants/diab022/6342117 by guest on 11 August 2021

1. Introduction

The circadian clock in plants temporally coordinates an extensive repertoire of developmental and physiological processes. These include seedling establishment, photosynthesis, cell division, and flowering time, amongst others. Its physiological relevance, architecture and response to environmental signals have been extensively reviewed (Millar 2016; Nohales and Kay 2016; Sanchez *et al.* 2020). The fact that mutations of the clock genes have such pleiotropic effects can be explained at the molecular level, by observations that more than 30% of the *Arabidopsis thaliana* transcriptome exhibits circadian rhythmicity (Covington *et al.* 2008; Edwards *et al.* 2006; Harmer *et al.* 2000). The seminal timeseries studies have been repeated in other species, demonstrating how the Arabidopsis results translate into crops (Edwards *et al.* 2018; Li *et al.* 2019; Matsuzaki *et al.* 2015; Müller *et al.* 2020, 2016). Studies of seasonal timing (phenology) in crops are also uncovering the effects of clock genes that are homologous to those first identified in Arabidopsis, and the impact of Genotype x Environment interactions (Bendix *et al.* 2015; Millar 2016). Understanding the clock gene circuit and its outputs to physiology and phenology therefore holds promise to guide further breeding and/or engineering of crop varieties (Preuss *et al.* 2012 p. 32) and the clock regulation of RNA levels is of particular interest.

Arabidopsis thaliana remains a critical, model organism for prototyping the theoretical and experimental tools that are subsequently combined with the methods from crop science. Systems biology approaches have long been applied to understand the complex mechanism of the Arabidopsis clock gene circuit, as noted in the reviews cited above, and introduced in detail below. However, even this paradigmatic system still falls short of the resources that are required to understand (explain and predict, quantitatively and mechanistically) how genome sequence variation at single base-pair resolution affects whole-organism traits (Millar 2016). This also hampers the selection of the genome sequence that is needed for some desired physiology (Marshall-Colon *et al.* 2017) whereas this possibility is already exemplified in other systems, such as the engineering of transcription factor binding sites in

Downloaded from https://academic.oup.com/insilicoplants/advance-article/doi/10.1093/insilicoplants/diab022/6342117 by guest on 11 August 2021

E. coli (Barnes *et al.* 2019). Aspects of the plant clock mechanism based on RNA metabolism have been more tractable than protein regulation to date, whether in RNA processing (James *et al.* 2012) or as here in transcriptional regulation (Kang *et al.* 2018; Lin *et al.* 2020; Xu *et al.* 2020). Even at this level, current models of plant gene regulatory networks represent their RNA components with arbitrary mass units, and hence their transcription and translation rates likewise. This practice is inherited from the experimental methodology, where RNA abundance is typically normalised to an internal standard for quantification (Czechowski *et al.* 2005), yielding data in arbitrary, relative units.

Refactoring the models to use absolute mass units would not only be a step towards engineering but should also allow more discriminating tests of model validity. First, any estimated parameter value should fall within a realistic distribution for the cognate biochemical process. Such a distribution is hard to generalise in arbitrary units. This test can already discriminate between models and reveal the system's evolutionary constraints, as shown for enzymatic reactions (Bar-Even *et al.* 2011). Collating biochemical parameter measurements to form an expected distribution is therefore beneficial (Jolley *et al.* 2014), similarly to the collation of trait data for physiology (Poorter *et al.* 2010), but has been rare for plant gene regulation (Millar *et al.* 2019). Here, we aim to contribute at this level. Second and more obviously, biochemical measurements should directly test the particular parameter values in the model, improving the model's realism over time.

This approach is most relevant to models that focus on biochemical fidelity (Fogelmark and Troein 2014; Locke *et al.* 2005b; a; Pokhilko *et al.* 2010, 2012, 2013; Zeilinger *et al.* 2006). Other clock models have used greater levels of abstraction to emphasise operating principles (De Caluwé *et al.* 2016, 2017; Foo *et al.* 2016, 2020) or to gain computational tractability when studying large ensembles (Greenwood *et al.* 2020). Our interest in linking genome sequence mechanistically to complex plant phenotypes requires significant

Downloaded from https://academic.oup.com/insilicoplants/advance-article/doi/10.1093/insilicoplants/diab022/6342117 by guest on 11 August 2021

biochemical and biophysical detail, so we consider two of the most detailed models available, referred to as P2011 (Pokhilko *et al.* 2012) and KF2014 (Fogelmark *et al.* 2014).

1.1 The clock mechanism represented in the models

Both models represent a set of interlocking, transcriptional, negative feedback loops, in which the clock genes control each other's expression at particular times of day, through their short-lived proteins (Figure 1a; Supplementary Table 1). 24-hour rhythms of gene activity with characteristic waveforms emerge from the gene network models' dynamics, without a single 'rhythm generator'. Clock gene expression is affected by several light input pathways, so the biological rhythms are entrained to the day/night cycle. Both models reflect the rhythmic expression profiles of clock components observed in several light conditions and mutant backgrounds with good fidelity, and in real time units.

Starting at dawn, expression of the highly-homologous LATE ELONGATED HYPOCOTYL (LHY) and CIRCADIAN CLOCK ASSOCIATED 1 (CCA1) genes peak simultaneously. The myb-related LHY and CCA1 proteins in the plant repress the *PSEUDO RESPONSE REGULATOR* gene family (*PRR9, PRR7, PRR5* and *TOC1* ≡ *PRR1*), and of GIGANTEA (GI), LUX-ARHYTHMO (LUX), EARLY FLOWERING 3 (ELF3) and EARLY FLOWERING 4 (ELF4) (Alabadí *et al.* 2001; Kamioka *et al.* 2016; Mizoguchi *et al.* 2002; Nagel *et al.* 2015; Nakamichi *et al.* 2010; Schaffer *et al.* 1998; Wang and Tobin 1998). The latter, 'evening genes' are repressed until LHY and CCA1 protein levels fall during the day, whereas the *PRR* genes are expressed in a sequence from the early morning through the day until after dusk (Fujiwara *et al.* 2008; Nakamichi *et al.* 2010), and repress the expression of *CCA1* and *LHY*. Matching the observed *PRR* dynamics in the precursor model of P2011 required a transcriptional activator function (Pokhilko *et al.* 2010). Some members of the *REVEILLE/LHY-CCA1-like* family of transcription factors, which are dawn-expressed genes homologous to *LHY* and *CCA1,* were later observed to function as transcriptional activators. Family members notably RVE8/LCL5 have been shown to interact with a family of

Downloaded from https://academic.oup.com/insilicoplants/advance-article/doi/10.1093/insilicoplants/diab022/6342117 by guest on 11 August 2021

transcriptional co-activators termed NIGHT LIGHT-INDUCIBLE AND CLOCK REGULATED GENES, the LNK1-LNK4 family (Rugnone *et al.* 2013), which were not represented in either model. The evening-expressed genes *LUX, ELF4* and *ELF3* together produce the hetero-trimeric protein Evening Complex (EC) (Nusinow *et al.* 2011), which binds to the promoters of *PRR9, TOC1/PRR1, LUX, ELF4* and *GI* and represses their expression, in turn allowing *LHY* and *CCA1* expression on the next cycle.

The P2011 and KF2014 models differ in their representation of several, relevant components and interactions (please see Supplementary Information). Our approach here is to re-factor the P2011 model to use absolute mass units, generating two alternative models, termed U2019 and U2020. The U2019 model is close to the P2011 model, whereas U2020 is revised to include key interactions from the KF2014 model in this simpler context. We use multiple data sources from the literature to test aspects of the models that were not previously accessible, in particular the models' inferred transcription rates.

2. Materials and Methods

All experimental data have been published elsewhere. The key results for model fitting derived from the TiMet project funded under the EU Framework Programme 7, and are publicly available as described (Flis *et al.* 2015). Other data sources are cited in the text. All model analysis was performed using python 2.7, in the computational environment described below and available as described in the Reproducibility section (section 2.4).

2.1 Model analysis and availability

The P2011 model was translated from MATLAB (Pokhilko *et al.* 2012) into the high-level Antimony language and then translated into SBML using the python package Tellurium (Choi *et al.* 2018). The model translations of P2011 are available in the public FAIRDOMHub repository at https://fairdomhub.org/assays/1219. The KF2014 model was originally written in clocksim (http://cbbp.thep.lu.se/activities/clocksim/F2014-20140425.tar.gz) (Fogelmark *et*

Downloaded from https://academic.oup.com/insilicoplants/advance-article/doi/10.1093/insilicoplants/diab022/6342117 by guest on 11 August 2021

al. 2014). As a contribution to the community, we also translated one of the parameterisations of the KF2014 model into Antimony and SBML, available at https://fairdomhub.org/studies/508. The U2019 and U2020 models were developed and analysed using Tellurium. Multiple versions of the model files including equations to generate light:dark cycles are available from the live data record at https://fairdomhub.org/investigations/170, which may include updated versions, and from the static "snapshot" (see section 6, Model and Data Availability). The minimal SBML files are directly accessible using the root path www.fairdomhub.org/models/. Each model can be accessed by appending the FAIRDOMHub ID listed below, for example U2019.1 can be accessed as www.fairdomhub.org/models/726, and similarly for U2019.2 (727), U2019.3 (728), U2020.1 (729), U2020.2 (730) and U2020.3 (731). Equations and parameter values were extracted for publication from these files using COPASI version 4.30.240 (please see Supplementary Information). SBML model files that include a standard light:dark function, the Input Signal Step Function (Adams *et al.* 2012) are included in the static snapshot of these and further resources for this publication (section 6, Model and Data Availability; Supplementary Information).

2.2 Model fitting

To estimate parameter values by comparison to experimental data, SBML models were imported into the SloppyCell python package (Myers *et al.* 2007). Step functions in SloppyCell were included to simulate diurnal changes of light conditions. Models were entrained for 10 cycles of 24 hours with 12 h of light and 12 hours of darkness, and the match to RNA timeseries data was tested using the χ^2 statistic, as follows. The TiMet data set for RNA profiles in one Light:Dark cycle (LD) followed by constant light was compared to simulations for the same conditions on simulation days 11 to 13, with constant light from dawn on day 12. To test the simulated rhythmic periods, the period of the cL_m variable (*CCA1/LHY* mRNA) was measured in simulated data starting from 276 h (12h after the start of constant light), except for the long-period *prr79* mutant which used data from 300 h, and

Downloaded from https://academic.oup.com/insilicoplants/advance-article/doi/10.1093/insilicoplants/diab022/6342117 by guest on 11 August 2021

the *cEC* variable was used for the *lhy/cca1* double mutant. The target period values for each genotype are described in Table 1.

To update the parameter values to better match experimental data, all the parameters in the model were allowed to vary except for the Hill coefficients and the set (*m11, m27, m31, m33, n14, n5, n6, p15, p6, p7*). This set contains parameters related to CONSTUTITIVE PHOTOMORPHOGENESIS 1 (COP1) which were inferred previously using data for ELONGATED HYOCOTYL 5 (HY5) and LONG HYPTOCOTYL IN FAR-RED (HFR1) proteins in separate experiments (Pokhilko *et al.* 2012), and parameters controlling the dynamics of the hypothetical protein P. Parameter sets were updated using the Levenberg-Marquardt method of SloppyCell. When fitting scaling factors, we used the sensitivity equation functionality built into SloppyCell, which updates efficiently by determining the direction of largest gradient change in the cost function landscape (Myers *et al.* 2007). For *ad hoc* constraints, such as period values and amplitude, SloppyCell provides prototype methods that were adapted to implement a gradient by finite difference approach. The TiMet timeseries data and period constraints were not sufficient to enforce oscillatory dynamics in constant light conditions, for the models described in this work, hence amplitude constraints were also included. The modified SloppyCell codebase was documented with git and is available from https://github.com/jurquiza/SloppyCell_Urquiza2019a. The modifications are highlighted in the code base with the string “Uriel”.

2.3 Stepwise development of model U2020.1 from U2019.1

To replace the *PRR* activation cascade in model U2019.1 with a repression cascade, we introduced a new set of variables that represent the *PRR* genes, with the revised regulation. The repressive function of CCA1/LHY (*cL*) and its activating derivative (cL_{mod}) were retained. The new equations did not initially feed back into the original equations, to avoid wholesale modification of the model’s behaviour. The parameters for these “dummy” equations were fitted to *PRR* profiles simulated from U2019.1 using the Levenberg-Marquardt algorithm of

Downloaded from https://academic.oup.com/insilicoplants/advance-article/doi/10.1093/insilicoplants/diab022/6342117 by guest on 11 August 2021

SloppyCell. The original *PRR* equations could then be replaced with the new equations, with only minor changes to model behaviour. The new model was then fitted to data simulated from U2019.1 for the entire set of variables, conditions and mutants present in the TiMet data set (leaving only parameters associated with COP1 and P fixed, as noted). The aim was to recover P2011 behaviour using the update network topology resulting in the model U2020.1

2.4 Reproducibility

We developed a Docker image (uurquiza/urquiza2019a_tellurium_sloppycell:latest) to ensure the reproducibility of computational results. The image is publicly available from the Docker hub (https://hub.docker.com) and from https://fairdomhub.org/assays/1224. After installing the Docker software, the image can be downloaded by typing docker pull image_name. The image can also be built from source code available in the git repository https://github.com/jurquiza/Urquiza2020a. This repository contains instructions for installation and for running the Docker image, which compiles all the required tools for model fitting to oscillatory dynamics. The fitting is described in Jupyter notebooks in the same git repository. These are linked from the resource snapshot for this publication (see section 6, Model and Data Availability).

3. Results

Before re-factoring the P2011 model to use absolute units, two aspects of this model were reconsidered in the light of recent evidence and insights from the KF2014 model (introduced in the Supplementary Information).

3.1 Transcriptional activation in the P2011 model and simulation of REVEILLE 8

We and others have focussed on the prevalence of negative regulation in the Arabidopsis clock circuit, and on transcriptional repression in particular. The P2011 model also includes transcriptional activation by the variable *Lmod* (denoted cL_{mod} in written model equations, or

Downloaded from https://academic.oup.com/insilicoplants/advance-article/doi/10.1093/insilicoplants/diab022/6342117 by guest on 11 August 2021

cLm in P2011 in SBML format). *Lmod* activates the expression of the later-peaking *PRR* genes *PRR7* and *PRR5* (Figure 1a; *PRR5* is identified with the Night Inhibitor predicted by the P2011 model). *Lmod* was first proposed in the earlier P2010 model, where this hypothetical component was required to match the extended profile of *PRR* gene expression (Pokhilko *et al.* 2010). It is represented as a modified version of the model's combined CCA1/LHY protein *L*. *Lmod* functions only as a transcriptional activator (Pokhilko *et al.* 2010), whereas the unmodified protein acts as the known repressor of evening genes, and to activate early *PRR* gene expression. CCA1 protein was known to be phosphorylated in a manner that altered DNA binding (Andronis *et al.* 2008), suggesting one potential biochemical mechanism for the modification. Slow conversion of *L* into *Lmod* leads to peak accumulation of *Lmod* at ZT 6.5 h (Figure 1b), providing a peak of activation at midday. The RNA variable that represents *CCA1* and *LHY* transcripts peaks just after dawn, in line with the experimental data. Therefore, the model includes a 6 h lag between the transcript peaking time and maximal accumulation of *Lmod*. This effective solution matched the observed expression profiles of the *PRR* genes and, given the lack of constraining evidence, it was presented as a parsimonious but phenomenological construct.

Interestingly, the REVEILLE 8 / LHY-CCA1-LIKE 5 (RVE8 / LCL5) gene product presents very similar behaviour (Rawat *et al.* 2011). The levels of *RVE8* transcript peak at dawn and RVE8 protein levels peak 6 h later (Figure 1b, RVE8-HA). In order to test whether *Lmod* in the model recapitulated the functions of *RVE8* in the plant, we sought to replicate two published perturbations of *RVE8*. The *rve8* mutant has a late rise of *CCA1* transcript under constant light conditions. Setting the transformation rate of *L* to *Lmod* (parameter *p3*) to zero in P2011 replicated the mutant phenotype (Figure 1c, *rve8*). This manipulation delays the repression of *CCA1* transcription and results in a later peak and long circadian period in constant light, also matching luciferase imaging data (Farinas and Mas 2011; Rawat *et al.* 2011). Experimental overexpression of *RVE8* (in *RVE8-OX* transgenic plants) results in period shortening (Farinas *et al.* 2011; Rawat *et al.* 2011; Shalit-Kaneh *et al.* 2018), whereas

Downloaded from https://academic.oup.com/insilicoplants/advance-article/doi/10.1093/insilicoplants/diab022/6342117 by guest on 11 August 2021

simulating constant high levels of *Lmod* resulted in arrhythmia. Though *Lmod* resembled *RVE8* in the loss-of-function test, it did not recapitulate the simplest simulation of *RVE8* overexpression.

We therefore tested whether an alternative mechanism could explain the behaviour of *RVE8-OX,* reasoning that RVE8 might function as a required component of a hypothetical Noon Complex (NC) of proteins, which collectively form the transcriptional activator. If other components of the complex are also rhythmically regulated with peak abundance around ZT6, then *RVE8* overexpression might accelerate NC formation without altering its peak time. Increasing the transformation rate of *L* into *Lmod* (parameter *p3*) simulated this effect, and recapitulated the early rise of *CCA1* transcript that is observed in the *RVE8-OX* genotype, along with its short period (Figure 1c) (Rawat *et al.* 2011).

Experimental data from studies of the LNK protein family (Rugnone *et al.* 2013) is consistent with the existence of a rhythmic Noon Complex. The *LNK* gene family is expressed in the morning and mutants of *LNK* genes present slower clocks, similar to the *RVE* family (De Leone *et al.* 2019; Hsu *et al.* 2013). Interaction between RVE and LNK proteins is reported *in vivo* (Xie *et al.* 2014) but plants that constitutively express both *RVE8* and *LNK1* do not show continuous interaction between these proteins. Rather, co-immunoprecipitation timeseries show that their interaction is still phase-specific (Pérez-García *et al.* 2015) (supplementary data). This result suggests that a third, rhythmically expressed component is required for complex formation (see Discussion, section 4.1), in line with the proposed Noon Complex. *Lmod* was proposed on functional grounds and might represent a functional Noon Complex more closely than it represents the individual, molecular constituent RVE8.

However, increasing the *L* -> *Lmod* transformation rate by increasing the parameter *p3* both increases *Lmod* and decreases *L* concentrations. Therefore, we dissected this dual effect (Figure 1d). Simulating only the increase of *Lmod* accumulation without impacting *L* levels,

Downloaded from https://academic.oup.com/insilicoplants/advance-article/doi/10.1093/insilicoplants/diab022/6342117 by guest on 11 August 2021

by introducing a dummy parameter that scales *p3* only for the *Lmod* synthesis equation, did not recapitulate the RVE8-OX phenotype, indicating that the post-translational, inhibitory effect on CCA1/LHY function is required in the model. This might represent RVE8 binding competitively with CCA1/LHY to DNA or inhibiting co-repressor recruitment by CCA1/LHY in the plant. We note that genetic evidence from the quintuple mutant *lhy/cca1/rve468* has been taken to suggest antagonism between the function of these RVE proteins and CCA1/LHY (Shalit-Kaneh *et al.* 2018). Analysis of Protein Binding Microarray profiles (Franco-Zorrilla *et al.* 2014) showed a significant correlation between the DNA binding motifs of CCA1 and RVE1 (no data are available for RVE4, 6 or 8), whereas the binding site preferences of CCA1 and LUX differed more in the same comparison (Supplementary Figure 1) even though LUX shares the Evening Element target sequence. This result is not surprising, as proteins that share DNA binding domain sequences also tend to have similar binding motifs (O'Malley *et al.* 2016; Weirauch *et al.* 2014). Nonetheless, our analysis supports the notion that competition for DNA binding might mediate the observed genetic antagonism, and our simulations suggest that this or another inhibitory mechanism contributes to the phenotype of RVE8-OX plants. The existing behaviour of *Lmod* concisely represents these effects. There is insufficient justification to complicate the model by representing the Noon Complex explicitly, with a time-constrained interaction of RVEs, LNK1 and an unknown, third component(s) (see Discussion section 4.1).

Finally, the formation of the Evening Complex in the model was revised to address a technical issue, relating to the form of the relevant equations (please see Supplementary Information; Supplementary Figure 2). The revision emphasised the role of the Evening Complex in the damped, short-period oscillations that remain in the *lhy;cca1* double mutant. The process for simulating this mutation was also revised, to remove only transcriptional repression by LHY/CCA rather than all the functions of both *L* and *Lmod* (see Supplementary Information). The baseline U2019 model for this study retained the arbitrary

Downloaded from https://academic.oup.com/insilicoplants/advance-article/doi/10.1093/insilicoplants/diab022/6342117 by guest on 11 August 2021

mass units as in P2011 but used the insights and updates outlined above, which should also be relevant to other clock models (Fogelmark *et al.* 2014; Pokhilko *et al.* 2013)

3.2 Development of the U2019.1 model

The updated model equations were initially fitted to simulation results from P2011 using the SloppyCell software (Myers *et al.* 2007), in order to recreate dynamics similar to P2011. The fitting process used simulated wild type plants (WT), and the mutants *lhy/cca1, prr7/9, toc1, gi* and *ztl*. The *ztl* mutant is important for constraining the decay rate of TOC1 protein, though it is absent from the TiMet RNA timeseries data used below. This first model takes the name U2019.1 (available from https://www.fairdomhub.org/models/726). Figure 2a outlines the nomenclature of subsequent model versions, while Figure 2b shows the contributing software and resources.

3.3 Structural changes for *PRR* regulation (backward temporal inhibition) in U2020.1

In P2011 the wave of PRRs is a consequence of mutual activation, in order CCA1/LHY (*Lmod*) -> PRR9 -> PRR7 -> PRR5 (Figure 1a). Experimental evidence suggests repressive activity for the PRRs (Nakamichi *et al.* 2012, 2010; Huang *et al.* 2012; Gendron *et al.* 2012; Liu *et al.* 2013), suggesting a mechanism in which a wave of inhibition CCA1/LHY (Lmod) |-- PRR9 |-- PRR7 |-- PRR5 |-- TOC1 can produce similar peak times (Figure 3a). This mechanism was partially introduced in the P2012 model, where TOC1 inhibits *PRR9* transcription (Pokhilko *et al.* 2013). Other modelling efforts expanded this idea to further PRRs (Fogelmark *et al.* 2014; Foo *et al.* 2016). Analysis of Foo et al. (2016) showed that repression results in cuspidate (sharp) peaks of gene expression, which could generate larger rhythmic amplitudes. High-amplitude regulation (100- to 1000-fold changes in RNA level) was a striking feature of the TiMet RNA timeseries, suggesting this mechanism could improve the model. Therefore, we updated the *PRR* activation chain to a repressive mechanism (Figure 3a), in a conservative, stepwise approach (see Methods section 2.3),

Downloaded from https://academic.oup.com/insilicoplants/advance-article/doi/10.1093/insilicoplants/diab022/6342117 by guest on 11 August 2021

ending with a broad re-optimisation of parameters to fit data simulated from U2019.1. Figure 3b shows simulated *PRR7* RNA timeseries from an intermediate stage. Through this approach, the repression-wave model U2020.1 recovered behaviours similar to the activation-wave model U2019.1, illustrated by the simulated *PRR7* transcript (Figure 3c), which also shows the short period expected in simulations of the *toc1* mutant. The U2020 model thus represents an intermediate between P2011 and the greater complexity of KF2014 (Fogelmark *et al.* 2014). With the U2019 and U2020 model circuits in hand (Supplementary Figure 3), we next develop a method to use transcript data in absolute units to recalibrate the plant clock models.

3.4 Data description

Flis et al (2015) performed absolute quantification of clock RNA transcripts, reporting data as average RNA copies/cell. The transcript collection consists of *LHY, CCA1, PRR9, PRR7, PRR5, TOC1, GI, LUX, ELF3, ELF4* and *GAPDH*. Timeseries studies that follow the emerging recommendations for quantification of these transcripts (Hughes *et al.* 2017) were performed in WT plants and four clock mutants (*lhy-21/cca1-11, prr9-11/prr7-11, toc1-101 and gi-201*). This provides a reference dataset for the transcriptional dynamics of plant clock models. Moreover, these data do not require internal normalising transcripts, reducing the risk of waveform distortion to the clock transcripts due to normalisation factors that might not remain constant over time.

3.5 The experimental error in calibrated qRT-PCR data

Biochemically-realistic models in molecular biology generally present a large number of parameters. Experimental approaches for determining these parameter values directly can be extremely challenging. A model-based approach can be taken for estimating parameter values from time-series data of observed variables, such as the RNA timeseries used to develop the earlier clock models, discussed above. If the experimental error is well

Downloaded from https://academic.oup.com/insilicoplants/advance-article/doi/10.1093/insilicoplants/diab022/6342117 by guest on 11 August 2021

characterised, then a maximum likelihood approach can be taken. In particular, if the experimental error follows a normal distribution then the following likelihood function can be used,

$$P\left(D\middle|M(\theta)\right) = \prod_{k=1}^{m}\prod_{i=1}^{d_k}\frac{1}{\sigma_{ki}\sqrt{2\pi}}exp\left\{-\frac{1}{2}\frac{\left(y_{ki}-y_k(t_i,\theta)\right)^2}{\sigma_{ki}^2}\right\} \quad (1)$$

This measures the probability of data *D* being generated by a model *M* with a parameter vector θ. Experimental measurements are represented by y_{ki} with uncertainty σ_{ki}. Model predictions for the measurements y_{ki} are given by $y_k(t_i,\theta)$. The *k* index denotes state variables, while the *i* index represents time points. A maximum likelihood estimate for the model parameters can be found by minimising the difference between model predictions and data (Figure 4a), and this general approach has been widely used. However, the nature of the experimental uncertainty ('error') is important in the inference process and this has been relatively neglected. The nature of the errors depends on the methodology used for gathering the experimental data (Raue *et al.* 2013).

Experimental error has been characterised in the context of the Arabidopsis clock, using RNA-seq data (Mombaerts *et al.* 2016). The TiMet data set used calibrated qRT-PCR data in units of [molecules] $[cell]^{-1}$, which need not have the same error distribution. Though the method avoids normalisation to an internal reference RNA, nonetheless Flis et al (2015) also quantified the transcript level for a common reference gene Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) using primers for both ends of the mRNA, in all the clock mutant backgrounds tested and across the 72 hours of sampling. These results for a nominally-constant RNA provide an opportunity to characterise the experimental error of this methodology.

Downloaded from https://academic.oup.com/insilicoplants/advance-article/doi/10.1093/insilicoplants/diab022/6342117 by guest on 11 August 2021

We compared the raw qRT-PCR data of GAPDH to the expected theoretical quantiles for a normal distribution. We observed departure from normality at both tails of the distribution, for both the 5' and 3' UTR of GAPDH (Figure 4b). It has been reported that experimental error in the determination of molecular levels of components can be described by log-normal models (Furusawa *et al.* 2005; Kreutz *et al.* 2007; Raue *et al.* 2013). We transformed the data by taking the natural logarithm, termed log-transformed data by convention. log-transformation of the data resulted in a better match to normality (Figure 4b). Therefore, data was log-transformed for evaluating model costs. Furthermore, thanks to the large number of data points reported by Flis et al ($n > 600$), we were able to estimate the associated uncertainty (5'-UTR, $\sigma = 0.95$; 3'-UTR, $\sigma = 0.34$), which might be expected in other data similar to the TiMet study. The uncertainties derived from biological replicates at each time point are preferred (n=2 in the TiMet data). Where a replicate is missing, the inferred uncertainty (using the average of the 5'-UTR and 3'-UTR values) provided a reasonable proxy. In case the value of the experimental error cannot be determined empirically, the researcher can attempt to infer the experimental noise during the parameter inference (Raue *et al.* 2013). A more extreme scenario is when the nature of the error function is unknown or difficult to determine. In this case an "error model average" likelihood approach could be taken (Kinney *et al.* 2010).

3.6 Period constraints

Circadian period estimates are the second data type used for model fitting, and these are available for the WT and for several mutants. It has been generally assumed that the error in period estimates follows a normal distribution. To test this, we analysed publicly available period estimates for transgenic Arabidopsis plants that report that transcriptional activity of two clock promoters, CCA1p:LUC and TOC1p:LUC, tested in longitudinal imaging assays. The data were deposited in Biodare (https://www.biodare.ed.ac.uk/, experiment ID 12730563219125). Testing for departure from normality results in non-significant measures both for CCA1p:LUC (statistic=1.01, p-value=0.6) and for TOC1p:LUC (statistic=0.078, p-

Downloaded from https://academic.oup.com/insilicoplants/advance-article/doi/10.1093/insilicoplants/diab022/6342117 by guest on 11 August 2021

value=0.67), (Figure 4c) (D'Agostino and Pearson 1973; D'Agostino 1971). Therefore, the period data are well represented by a normal distribution, with σ = 0.52 for CCA1p:LUC and σ = 0.71 for TOC1p:LUC. Our period constraints derived from imaging studies using CCA1p:LUC, so the cognate standard deviation was incorporated into the likelihood function.

3.7 Linking U2019.1 and U2020.1 to TiMet time-series and period constraints.

With confidence in our normality assumptions, we next linked the models to the TiMet data set. First, we introduced mass scaling factors for each transcript variable. These result only in a change in the units of RNA variables, which is obtained by multiplying each transcription rate by its scaling factor and dividing the cognate translation rate by the same factor. We fitted the scaling factors of U2019.1 and U2020.1 to log-transformed TiMet RNA timeseries data using SloppyCell. The resulting, rescaled models are named U2019.2 and U2020.2, and have units of [transcripts][cell]$^{-1}$ for the transcript variables (Supplementary Figure 4). Up to this point, however, the dynamic behaviour of the models was constrained by the behaviour of the P2011 model, not by the experimental timeseries.

The rescaled models were then fitted to the TiMet RNA timeseries for WT, *lhy/cca1, prr7/9, toc1* and *gi* mutants, and to the corresponding period estimates of transgenic plants with these genotypes in constant light (Table 1). The period of the *ztl* mutant was again included to constrain the degradation rates of TOC1 protein. The resulting models are named U2019.3 and U2020.3, derived from U2019.2 and U2020.2 respectively. Figure 5 shows a sample of the fitted time-series, from simulations which correspond to Col-0 WT data for *LHY*, *PRR7, TOC1* and *LUX*. Table 1 shows that both models matched the target periods very closely. Furthermore, both models corrected the phase delay that the P2011 model showed upon transfer from Light:Dark (LD) cycles to constant light (compare Figure 5 to Figure 1d), when the model's rapid degradation of PRR proteins in darkness is lost (Flis *et al.* 2015). The repression-based model U2020.3 shows a higher amplitude of RNA regulation, closer to the data for *LHY*, *PRR7* and *LUX* (see Discussion section 4.3). Overall,

Downloaded from https://academic.oup.com/insilicoplants/advance-article/doi/10.1093/insilicoplants/diab022/6342117 by guest on 11 August 2021

we observe a slightly better fit to the data for U2019.3 (total cost $6.16x10^6$) compared to U2020.3 (total cost $1.02x10^7$), and U2019.3 has slightly fewer parameters (99 compared to 103, Supplementary Table 1). Updating the regulatory mechanisms did not guarantee an improved fit to the data.

Table 2 shows the cost for each individual RNA timeseries in U2019.3 compared to U2020.3, highlighting which aspects of model behaviour were affected by the change in *PRR* gene regulation. The U2020.3 simulation of the Col-0 WT returned a better match (lower cost value) than U2019.3 for the genes *TOC1, PRR7, LHY, LUX* and *GI*. However, the absolute cost of *PRR9*, *LHY* and *GI* is higher for both models than for example *LUX* in U2020.3, where the simulated expression profile presents a notably good match to the high-amplitude regulation in the TiMet data (Figure 5; Supplementary Figure 6).

Compared to the data for clock mutants, simulation of the *lhy/cca1* double mutant improved the fit for several variables in U2020.3, with moderately increased costs for *PRR7* and *PRR5*. The *gi* mutant data was also better matched by U2020.3, with a particular improvement in the simulation of *PRR5*. U2019.3 matched better than U2020.3 to the data for the *prr7/9* mutant, where the match of U2020.3 to data for *PRR5* expression had the greatest cost of all the timeseries. At the other end of the *PRR* cascade, the *toc1* mutant data was also better matched by the U2019.3 model, with *PRR9, PRR5* and *LUX* contributing to higher costs for U2020.3. Across all the genotypes, the regulation of these three transcripts would merit further attention, based on the fitting data from U2020.3.

3.8 A genome-wide transcription rate distribution in absolute units

The recalibrated models now include transcription rates in absolute units that are compatible with the TiMet data and period constraints. The values of these model parameters (or strictly, the rates that are realised in simulation) can be taken as predictions of the functional rates *in vivo*. To provide a first-order test of these predictions, we constructed an empirical

Downloaded from https://academic.oup.com/insilicoplants/advance-article/doi/10.1093/insilicoplants/diab022/6342117 by guest on 11 August 2021

probability distribution of Arabidopsis transcription rates in units of [transcripts][cell]$^{-1}$[h]$^{-1}$, matching the model transcription rate parameters. No such data existed but two published data sets could be linked to derive the distribution, using simple assumptions.

The first data set consists of microarray transcriptomic measurements of synthesis and decay of RNAs in Arabidopsis, obtained in the BBSRC ROBuST project using the labelling kinetics of RNA with the ribonucleotide analogue 4SU (Sidaway-Lee *et al.* 2014), hereafter termed the Sidaway-Lee data (Figure 6b). This data includes only 7,291 genes which represents 35% of the 20,568 genes in the Arabidopsis genome (BNID 105446 https://bionumbers.hms.harvard.edu). The distribution has units of [microarray units] [cell]$^{-1}$[h]$^{-1}$. In order to transform the units of the distribution into [transcripts][cell]$^{-1}$ [h]$^{-1}$, we calibrated the Sidaway-Lee data using the second published data. The 'Piques data' comprises measurements at dawn and dusk of 96 transcripts coding for enzymes of Arabidopsis central metabolism (Piques *et al.* 2009), from the same, calibrated qRT-PCR method as (Flis Anna *et al.* 2015). We selected a subset of nine transcripts that were measured in both data sets and that changed less than 0.2-fold in level between dawn and dusk in the Piques data (Figure 6a). Assuming that no dynamic regulation of their transcription rates was taking place in either study, the data should be equivalent.

The recalibration starts from a simple, dynamic model for the production of each transcript,

$$\frac{dM_s}{dt} = a_s - bM_s \quad (1)$$

where M_s stands for the transcript level, a_s stands for the transcription rate and b for the transcript decay rate, and the subindex s refers to the Sidaway-Lee set. The units in this

Downloaded from https://academic.oup.com/insilicoplants/advance-article/doi/10.1093/insilicoplants/diab022/6342117 by guest on 11 August 2021

equation are [microarray units] $[h]^{-1}$. We can model a quasi-steady state by setting the differential equation to zero and solving for a_s,

$$a_s = bM_s \quad (2)$$

For the Piques data set the same model can be postulated, using parameter subindex *p*, but the transcription rates a_p have units of [transcripts]$[cell]^{-1}[h]^{-1}$. The decay constant *b* of each transcript should be equivalent in the two data sets, with units $[h]^{-1}$, so we omit the subindices. The link between the two data sets can be achieved by performing a linear regression of Sidaway-Lee vs Piques data sets. The resulting regression model is substituted in the Piques quasi steady state equation as follows,

$$a_p = bM_p \quad (3)$$

$$a_p = b(gM_s + i) \quad (4)$$

The gradient *g* has units of [transcripts][microarray units]$^{-1}$. We can now compare Sidaway-Lee data in [microarray units] $[h]^{-1}$ on the right-hand side of the equation with Piques data in [transcripts]$[cell]^{-1}[h]^{-1}$ on the left-hand side. Figure 6c shows a highly significant correlation between the data sets (r^2 ~ 0.8), and Figure 6d applies the regression to rescale the Sidaway-Lee distribution, yielding an average transcription rate of 38.4 in units of [transcripts]$[cell]^{-1}[h]^{-1}$, with a one standard deviation span from 12.42-119.7 [transcripts]$[cell]^{-1}[h]^{-1}$. This distribution should significantly aid the assessment of transcription rates in plant gene network models. We refer to it below as the Plant Empirical Transcription Rate distribution (PETR, pronounced Peter).

The linear regression approach described above can be tested using data sets that we

Downloaded from https://academic.oup.com/insilicoplants/advance-article/doi/10.1093/insilicoplants/diab022/6342117 by guest on 11 August 2021

expect to deviate more from our assumptions, where the correlation r^2 should drop. Fortunately, Sidaway-Lee et al described microarray studies for plants grown at two different temperatures 17 ºC and 27 ºC. The 17 ºC data is used above as those conditions are closer to the conditions used by Piques et al. Encouragingly, the 27 ºC data for the same genes returns a lower correlation coefficient (r^2 = 0.4) (Supplementary Figure 5, compare to Figure 6c). This result gives some support for the regression model used to develop the PETR distribution.

3.9 Challenging predicted transcription rates with the PETR distribution

We tested the fitted and scaled clock models U2019.3 and U2020.3 by comparing the maximum transcription rates reached in simulations to the PETR distribution (Figure 7a, 7b). In both models, the transcription rates fall entirely within the distribution, with few in the tails. cG_m transcription has a relatively high maximum (fast rate) compared to the other clock genes, for both models. This could reflect the requirement for strong, acute light activation of *GI* at dawn. In the case of U2019.3, CCA1/LHY presents a similarly fast maximum transcription rate, which is lower in U2020.3. For U2020.3, *PRR7* presents a rate among the highest estimates in the PETR distribution. These results indicate that the model circuits developed without the constraints of absolute mass units nonetheless predicted biochemically feasible transcription rates, potentially due to constraints on the cognate RNA degradation rates.

3.10 Testing the RNA degradation rates of models against the Sidaway-Lee distribution

The 4SU-labelling approach also estimated RNA degradation rates (Sidaway-Lee *et al.* 2014), so the distribution of these observed rates can likewise test the inferred decay rate constants in the U2019.3 and U2020.3 models (Figure 7c, 7d). Again, most decay constants fall within the realistic range for Arabidopsis. The clock-associated genes tend to have transcripts with a short half-life, consistent with their dynamic regulation. The *LHY* RNA

Downloaded from https://academic.oup.com/insilicoplants/advance-article/doi/10.1093/insilicoplants/diab022/6342117 by guest on 11 August 2021

degradation rate in the light (cL_mL) in U2020.3 is at the lower end of the distribution, whereas the rate in darkness (cL_mD) is at the higher end. In the case of U2019.3, the *PRR9* RNA degradation rate is higher than any observed value. The *GI* RNA degradation rate in both U2019.3 and U2020.3 models also falls clearly outside the distribution and is likely to be unrealistically high. The *PRR5* RNA in the U2020.3 model presents the most extreme value of all the transcript degradation rates, and is not biochemically realistic. We offer further interpretation of these results in the Discussion.

4. Discussion

Quantitative modelling of the Arabidopsis circadian clock mechanism has progressed significantly over more than 15 years (Bujdoso and Davis 2013). The complexity of the models has increased along with the inclusion of experimentally-documented variables (Supplementary Table 1). Naturally, most of the work has concentrated on simulating the timing of molecular events. The models have been challenged most stringently using timeseries data from mutant plants that alter rhythmic period and/or the waveform of clock components. Few of the inferred biochemical parameter values in the models could be tested against directly measured values, because neither the model parameters nor most of the available data were calibrated to absolute mass units that would allow direct comparisons. This work offers that calibration, for RNA levels and transcription rates. Our results derived from applying data science and modelling approaches to the outputs of three linked, systems biology funding awards (see Supplementary Information; Supplementary Figure 7).

4.1 Interpretation of clock models

Some of the revisions of the P2011 model equations in the new U2019 and U2020 models were required to avoid earlier assumptions in the formation of the Evening Complex and had little effect on model dynamics, but unfortunately increased model complexity from 28 to 31 variables. In contrast, recognising the role of transcriptional activation has altered how we interpret and use the models, without increasing their complexity. In the P2011 model, the

Downloaded from https://academic.oup.com/insilicoplants/advance-article/doi/10.1093/insilicoplants/diab022/6342117 by guest on 11 August 2021

activation required for clock gene expression during the day was captured by a combination of CCA1/LHY protein and a modified version termed *Lmod*, which had to form slowly in the model to activate *PRR* gene expression later in the day. In particular, *Lmod* regulates *PRR5* in the models (Pokhilko *et al.* 2012). Given the later discovery and characterisation of the RVE and LNK families of transcription factors, we note the similarity between *Lmod* and RVE8 / LCL5, particularly with respect to an enigmatic delay in observed RVE8 protein accumulation that matches *Lmod*. *PRR5* is also a key, direct target of RVE8 (Farinas *et al.* 2011; Hsu *et al.* 2013; Rawat *et al.* 2011; Shalit-Kaneh *et al.* 2018; Xie *et al.* 2014). The activating function in our models was little discussed (Somers 2012), in part because it pre-dated the most relevant data.

Both the U2019 and U2020 models matched well to the target period values and RNA timeseries in wild-type and mutant plants (Tables 1 and 2; Supplementary Figure 6), though the models differ in the mechanisms of *PRR* gene regulation (Figures 1a and 2a). The U2020 model had a slightly higher fitting cost and larger parameter number, so U2019 would be preferred *a priori*. However, the regulatory interactions in U2020, derived from the KF2014 model, are better evidenced. As the number of plant clock models grows, it will be increasingly important to compare and document their performance in detail. If this is a collaborative effort between the model developers, it may also reveal implicit/tacit knowledge (Davies 2001) that is codified in each model, and which is then applied to simulate specific biological scenarios, exemplified by our reinterpretation of the *lhy/cca1* double mutant.

Identifying *Lmod* as a model of the observed, daytime transcriptional activator has a technical consequence, that simulations of the *lhy/cca1* double mutant in several models (P2010, P2011, P2012, F2016) should not remove all CCA1/LHY protein but instead should selectively remove its repressive effects. Simulating the damped, short-period rhythms in this mutant has been technically challenging (Zeilinger *et al.* 2006). The model retains very few feedback loops in this case, so its behaviour is highly sensitive to some of the parameters

Downloaded from https://academic.oup.com/insilicoplants/advance-article/doi/10.1093/insilicoplants/diab022/6342117 by guest on 11 August 2021

describing their biochemistry. The challenge remains when activating and repressing functions are represented as separate genes (Fogelmark *et al.* 2014) or as effects on separate targets (this work; Flis et al., 2015). High parameter sensitivity is a marker for caution in modelling (Morohashi *et al.* 2002). The *lhy/cca1* double mutant simulations probably highlight the need for future work, which might have implications for our current view of the clock mechanism beyond this special case.

The biochemistry of transcriptional regulation contributes to the modelling challenges. Simulation of the clock dynamics in *RVE8-OX* plants encouraged us to propose the presence of a Noon Complex, formed by RVE8 and at least one other rhythmic partner, which might correspond more closely to the function of *Lmod* in the models. The LNK proteins are possible partners that interact with RVE proteins (Xie *et al.* 2014). However, in plant lines that constitutively express both RVE8 and LNK1, their interaction is still restricted to around ZT7 (Pérez-García *et al.* 2015), consistent with the requirement for at least one further, rhythmic partner. Affinity purification and mass spectrometry (AP-MS) could help in elucidating the cryptic component(s) required for NC formation and functionality, as applied to the Evening Complex (Huang *et al.* 2016) or in timeseries of interactions with the GI protein (Krahmer *et al.* 2019). Adding more of these processes explicitly will not only further complicate the models but also increase the uncertainty of their outputs, unless the new components are constrained by significant, additional data.

4.2 PETR, a constraining distribution of transcription capacity for plant systems biology

By recalibrating our models into absolute units of transcripts per cell using the TiMet qRT-PCR timeseries, we made it possible to test the modelled transcription rates against biochemical data, if such data existed. Though we previously collated distributions of the binding affinities measured for plant transcription factors to DNA and for protein-protein interactions in Arabidopsis (Millar *et al.* 2019), we had relied on data from yeast to illustrate

Downloaded from https://academic.oup.com/insilicoplants/advance-article/doi/10.1093/insilicoplants/diab022/6342117 by guest on 11 August 2021

the possible transcription rates in plant cells. Integrating two apparently different data sets here resulted in the PETR distribution (Figure 6d), which estimates a mean transcription rate of 38.4 [transcripts][cell]$^{-1}$[h]$^{-1}$, with one standard deviation spanning 12.4-119.7 [transcripts][cell]$^{-1}$[h]$^{-1}$. Yeast for comparison has rates of 2 - 30 [transcripts][h]$^{-1}$ (Pelechano *et al.* 2010). The higher mean in plants likely relates in part to the difference in size between yeast (43 μm^3) and Arabidopsis cells (73,000 μm^3) (Jorgensen *et al.* 2002; Wuyts *et al.* 2010). If the parameter values of other models are refactored into absolute units, the PETR distribution can be used to assess the resulting transcription rates *post hoc*. More likely, extensions to Bayesian inference, previously applied to the P2011 model (Higham and Husmeier 2013), might use the PETR distribution as a prior to constrain parameters to plausible values during model development. Directly measuring biochemical parameter values, including the absolute transcription rates of the clock genes, is still an important step for future model development.

The PETR distribution was derived as a log-normal distribution, following our assessment of the distribution of GAPDH qRT-PCR results in the TiMet data (Figure 4). Log-normal distributions are pervasive in biology, for example in western blot data (Kreutz *et al.* 2007). The statistical assumptions in parameter inference methods require a good description of the experimental noise, but it can be unclear what error distribution is appropriate, for example if time series have had to be extracted from data displays in the literature (Fogelmark *et al.* 2014). This underlines again the importance of routinely sharing the numerical values of biological data, in line with FAIR principles (Findable, Accessible, Interoperable, Re-usable). As FAIR applies to secondary as well as primary data, we have registered the key values of the PETR distribution in the Bionumbers database (see Model and Data Availability, section 6) (Milo *et al.* 2010).

Downloaded from https://academic.oup.com/insilicoplants/advance-article/doi/10.1093/insilicoplants/diab022/6342117 by guest on 11 August 2021

4.3 The implications of inferred clock gene transcription rates, and future model development

The key concerns in the parameter values governing the model's RNA metabolism arose from the high RNA degradation rates predicted for some clock transcripts (Figure 7c, 7d). These reflect the timeseries data that the model fitting seeks to match, within the constraints of the model equations. Measured *GI* mRNA levels fall over 1100-fold in the 12 hours after dusk in the TiMet data, for example (Flis *et al.* 2015). The *GI* mRNA in the model has a constant half-life, which must be short enough to match these dramatic decreases in mRNA level. However, *GI* mRNA also accumulated 32-fold in the 2 hours after dawn (Flis *et al.* 2015), and even more rapid increases have been observed (Locke *et al.* 2005). Rapid accumulation of this very unstable transcript in turn requires a high, estimated *GI* transcription rate in the models (Figure 7a, 7b), which might be reduced if the degradation rate of the *GI* transcript was also regulated. For transcripts of *CCA1/LHY*, in contrast, the half-life is not constant as the model includes their observed, light-regulated degradation rate (Yakir *et al.* 2007). De-stabilisation of this mRNA in the dark helps to reach lower, trough levels in our model simulations (Figure 5) than in P2011 and earlier models (Figure 1b), more closely matching the data. Comparing our two models, the *CCA1/LHY* transcription rate is unusually high in the U2019.3 model but less so in U2020.3. One reason for this is likely that the difference between light and dark mRNA degradation rates is greater for the cognate RNA in U2020.3, in other words, the RNA is much less rapidly degraded in the light compared to darkness, lowering the demand on new transcription to match the observed RNA levels (Figure 7). Unusually high estimates of RNA metabolism parameters for *PRR* transcripts such as *PRR5* might also suggest that these RNAs are post-transcriptionally regulated *in vivo*, as is common in other clock systems (Martelot *et al.* 2012).

PRR5 RNA expression not only has a high amplitude but also a cuspidate (peaked) waveform. Both are indicative of highly non-linear regulation, which might involve both regulated RNA stability and high-amplitude transcriptional regulation. The earliest modelling

Downloaded from https://academic.oup.com/insilicoplants/advance-article/doi/10.1093/insilicoplants/diab022/6342117 by guest on 11 August 2021

of circadian clocks used large Hill coefficients, which were later considered unrealistic, to simulate the cooperativity of transcriptional repressors. Our models conservatively limit the non-linearity of transcription factor binding, using Hill coefficients fixed at 2, which models the presence of a functional dimer. However, some clock promoters have multiple, similar DNA sequence elements that might bind several dimers, and competition between alternative binding partners at a single binding site can also generate larger apparent Hill coefficients (Buchler and Louis 2008; Kim and Forger 2012; Lee and Maheshri 2012). Understanding such competition, for example among the RVEs and other transcription factors that bind the Evening Elements, will require the absolute quantitation of the copy number of regulatory proteins per cell, the number of genomic binding locations and the transcription factor affinity for those locations. Protein quantification is therefore another critical step for plant clock models. The low abundance of clock proteins makes this challenging, and specific antibodies are available for only a few of the relevant proteins. Fusion proteins tagged using either firefly luciferase or NanoLUC offer a convenient alternative approach to direct protein quantification (Urquiza-García and Millar 2019). Indirect methods might also contribute to this challenge, such as the recently-described rhythmic translatome (Bonnot and Nagel 2021).

Future digital organism models will require both simplified but still experimentally-based models that facilitate analysis, and also biochemically-realistic models that link to genome sequences. These, detailed models should help to mobilise contributions from biochemistry, structural and chemical biology, which have influenced work on mammalian and cyanobacterial clocks far more than plant chronobiology. The detailed models will also assist the model-driven design of plant regulation, using precision genetic technologies such as Prime Editing (Lin *et al.* 2020).

Downloaded from https://academic.oup.com/insilicoplants/advance-article/doi/10.1093/insilicoplants/diab022/6342117 by guest on 11 August 2021

5. Model and Data Availability

The data, models, Docker containers and other resources are shared collectively as a static Snapshot, formatted as a Research Object and structured according to the standard ISA hierarchy, on FAIRDOMHub.org (doi: 10.15490/FAIRDOMHUB.1.INVESTIGATION.170.2), on the Zenodo repository (doi: 10.5281/zenodo.5150793) and on Edinburgh DataShare (doi: ??to be added in proof??). The data should be cited as:

Urquiza Garcia, U. & Millar, A. J. (2021). Absolute units in Arabidopsis clock models up to U2020.3 [Data set]. Zenodo. https://doi.org/10.5281/ZENODO.5150793

The live, updatable resources for individual models are also public on FAIRDOMHub.org, as described above in section 2.1, Model analysis and availability. The TiMet RNA timeseries data were already publicly available as experimental replicate data from the BioDare repository (experiment ID 2814; https://www.biodare.ed.ac.uk/robust/ShowExperiment.action?experimentId=2841). The live resource on FAIRDOMHub.org presents separate files of mean and standard deviation values, as used in this work (Data file IDs 2016 (means) and 2017 (SD), https://fairdomhub.org/data_files/2016 and https://fairdomhub.org/data_files/2017). The values of the distributions inferred above have been registered with the Bionumbers database as follows:

mRNA degradation rate distribution for *Arabidopsis thaliana*, mean ID??to be added in proof?? and SD ID??to be added in proof??.

PETR distribution, mean ID??to be added in proof?? and SD ID??to be added in proof??.

Downloaded from https://academic.oup.com/insilicoplants/advance-article/doi/10.1093/insilicoplants/diab022/6342117 by guest on 11 August 2021

6. Tables

Table 1. Period constraints enforced in model development.

The target periods (Enforced) and the observed periods for the U2019.3 and U2020.3 models are listed, for the Col-0 wild type and the clock mutants tested by simulation. The set of target periods is retained from earlier model development (Flis *et al.* 2015).

Genotype	Enforced (h)	U2019.3 (h)	U2020.3 (h)
Col-0	24.5	24.5	24.48
prr7/9	30	30.01	29.99
lhy/cca1	17	16.7	16.99
gi	22	22	22
toc1	21	20.99	20.98
ztl	27	27.01	27

Downloaded from https://academic.oup.com/insilicoplants/advance-article/doi/10.1093/insilicoplants/diab022/6342117 by guest on 11 August 2021

Table 2. Cost function values of the U2019.3 and U2020.3 models compared to the RNA time series.

Higher χ^2 values reflect a greater departure of each simulated RNA variable from the cognate TiMet RNA data, under one LD cycle followed by LL. A higher Ratio between these values reflects a greater departure of the U2020.3 simulation from the data, compared to U2019.3. Variable names are listed as the SBML variable identifiers; the cognate RNA data are as follows: cT_m, *TOC1*; cP7_m, *PRR7*; cP9_m, *PRR9*; cP5_m, *PRR5*; cE3_m, *ELF3*; cE4_m, *ELF4*; cL_m, *LHY*; cLUX_m; cG_m, *GI*. Some RNA species are absent in the mutants. The 'log' prefix denotes the log-transformation of both data and simulated values. Ratios above 1 are bold-faced and highlighted in red

		model χ^2		
Genotype	Variable	U2019.3	U2020.3	Ratio
WT (Col-0)				
	log_cT_m	6616.39	1872.18	0.28
	log_cP7_m	3049.33	1945.31	0.64
	log_cP9_m	1096000.94	2360519.61	**2.15**
	log_cP5_m	7154.91	57504.36	**8.04**
	log_cE3_m	11024.96	13146.04	**1.19**
	log_cE4_m	1086.76	1279.85	**1.18**
	log_cL_m	720144.18	161622.64	0.22
	log_cLUX_m	1951.47	743.42	0.38
	log_cG_m	100108.26	73998.62	0.74
gi				

Downloaded from https://academic.oup.com/insilicoplants/advance-article/doi/10.1093/insilicoplants/diab022/6342117 by guest on 11 August 2021

	log_cT_m	447.70	2126.95	**4.75**
	log_cP7_m	467650.36	343409.45	0.73
	log_cP9_m	73798.39	112937.04	**1.53**
	log_cP5_m	1162418.53	429889.14	0.37
	log_cE3_m	692.91	1857.66	**2.68**
	log_cE4_m	9729.48	9791.66	**1.01**
	log_cL_m	840604.60	771845.84	0.92
	log_cLUX_m	4678.02	4126.69	0.88
lhy/cca1				
	log_cT_m	3941.60	1563.78	0.40
	log_cP7_m	41669.61	94306.14	**2.26**
	log_cP9_m	8724.69	3446.11	0.39
	log_cP5_m	351321.63	702794.35	**2.00**
	log_cE3_m	835.21	285.74	0.34
	log_cE4_m	18568.11	20089.85	**1.08**
	log_cLUX_m	22099.10	7347.42	0.33
	log_cG_m	31712.52	31187.09	0.98
toc1				
	log_cP7_m	26696.91	43231.76	**1.62**
	log_cP9_m	17762.85	517633.15	**29.14**
	log_cP5_m	29305.91	274211.13	**9.36**
	log_cE3_m	1236.11	3393.01	**2.74**
	log_cE4_m	1228.35	1653.03	**1.35**
	log_cL_m	31066.85	11209.69	0.36
	log_cLUX_m	415452.95	1005765.19	**2.42**
	log_cG_m	25564.76	18017.81	0.70
prr7/9				
	log_cT_m	3498.73	722.99	0.21

Downloaded from https://academic.oup.com/insilicoplants/advance-article/doi/10.1093/insilicoplants/diab022/6342117 by guest on 11 August 2021

	log_cP5_m	549882.29	3063245.89	**5.57**
	log_cE3_m	2221.08	7760.54	**3.49**
	log_cE4_m	13908.81	9930.81	0.71
	log_cL_m	26982.44	17502.82	0.65
	log_cLUX_m	24255.14	13117.07	0.54
	log_cG_m	8003.18	5281.46	0.66
Total Chi^2		6163096.03	10202313.29	1.66

Downloaded from https://academic.oup.com/insilicoplants/advance-article/doi/10.1093/insilicoplants/diab022/6342117 by guest on 11 August 2021

7. Figure Legends

Figure 1

Transcriptional activation in P2011, compared to expression and function of *RVE8*.

a) Simplified diagram locating *Lmod* in the P2011 model. Arrows, activation; blunt lines (T), repression. Solid lines, transcriptional regulation; dashed red lines, posttranslational regulation. Yellow circles, light regulation. Genes are represented by round boxes. Relevant protein species are represented by circles. The equivalent regulation of *CCA1/LHY* by each PRR protein is indicated by the dashed black box. The proposed Noon Complex (NC) formed from RVE, LNK and at least a third unknown component is represented by a pie. b) P2011 simulation in L:D cycles (white:dark grey bars) showing the behaviour of CCA1/LHY variables cL_m (mRNA), *cL* (protein), cL_{mod} (modified, activating protein) and the similarity of *cLmod* to RVE8 protein data (RVE8-HA, (Hsu *et al.* 2013)). Peak expression times are indicated: 0.6 h cL_m, 2.6 h *cL* and 6.5 h c*Lmod*. c) Simulated perturbations in P2011 in Arbitrary Units (A.U.) that reproduce the rhythmic phenotypes of Col-0 (WT), *rve8* and *RVE8-OX* plants. The model was entrained for 10 days in L:D cycles (white:dark grey bars) before transfer to constant light (white:light grey) at time 0. d) Inhibiting *L* function is a necessary part of *Lmod* function. *RVE8-OX* was simulated by increasing the model parameter *p3* (as in c), purple line) including the inhibitory effect of *Lmod* formation that partly consumes *L*. In contrast, the *RVE8-OX* non-inhibitory simulation (non-inh., yellow line) generated the same level of *Lmod* without depleting *L*, and failed to match the *RVE8-OX* phenotype. Simulations were otherwise conducted as in c) with Arbitraty Units (A.U.).

Figure 2

Model development with open and reproducible resources.

a) Model nomenclature. Model architectures are denoted by author surname and year, with versions or parameter sets after a decimal point(s) (Flis *et al.* 2015). The published P2011.1 model simulated perfect data (blue). The U2019.1 model derived from P2011.1, without its

Downloaded from https://academic.oup.com/insilicoplants/advance-article/doi/10.1093/insilicoplants/diab022/6342117 by guest on 11 August 2021

quasi-steady state assumptions. U2020.1 has the *PRR* repression cascade (Figure 3 and main text). Parameters of both U2019.1 and U2020.1 were fitted to the perfect data of P2011.1. Models U2019.2 and U2020.2 were obtained by rescaling only transcription and translation rates such that modelled RNA levels matched TiMet RNA data (orange). The U2019.3 and U2020.3 models resulted from a near-global re-optimisation to TiMet RNA timeseries, period and amplitude constraints (green). b) Reproducibility and wide reuse of models requires open data analysis tools. Public reference data of the TiMet project and other literature resources informed model development in open languages such as Python and Jupyter. The SBML model exchange format is supported by an ecosystem of open-source software for systems biology, such that models can be written in Antimony, explored in Tellurium, and fitted in SloppyCell. Models can then be transferred to Tellurium for further analysis. FAIRDOMHub and GitHub are crucial for accessibility and version control. We use Docker and DockerHub to disseminate our software environment reproducibly.

Figure 3

The refactored, repression cascade of *PRR*s in the U2019 and U2020 models.

a) Simplified diagram of the *CCA1/LHY* and *PRR* gene families in the U2020 model, using the same conventions as Figure 1a. The Noon Complex (NC) proposed to function similarly to *Lmod* is shown as a pie, potentially formed from RVE, LNK and at least a third unknown component. b) Dynamics of *PRR7* transcripts in dummy variables controlled by U2019.1 under LD cycles then constant light; original, U2019.1 activation model, as in P2011 (dashed blue line); proposed, dummy variable for repression model (solid pink line). c) Dynamics of *PRR7* transcript in U2019.1 (dashed lines) and U2020.1 (solid lines), in WT (blue) and *toc1* mutant (yellow). The derivation of U2020.1 is described in the main text. The models were entrained for 10 days in L:D cycles (white:dark grey bars) before transfer to constant light (white:light grey) at time 0.

Downloaded from https://academic.oup.com/insilicoplants/advance-article/doi/10.1093/insilicoplants/diab022/6342117 by guest on 11 August 2021

Figure 4

The cost function for model development uses experimental error estimates.

a) The function used to calculate the cost $C(\theta)$, for a parameter set θ compared to data from one genotype, sums costs for all datapoints i, comprising the departure of log-transformed, scaled by analytically derived B_i scaling factors, simulated variable $y_i(\theta)$ from the cognate, log-transformed RNA timepoint d_i, scaled by the experimental error estimate σ of that data point, and the departure of simulated period τ_i from observed period τ, scaled by the experimental error estimate σ_p of that period value and N the number of independent period estimates; here, $N = 1$. B_i scaling factor for b) and c), the appropriate data transformation is inferred by the (mis)match between the distribution of a large set of (b) RNA or (c) period measurements and the expectation from the normal distribution (lines). b) quantile-quantile plots of GAPDH 5' (left panels) or GAPDH 3' RNA (right) data for untransformed (upper) and log-transformed data (lower), with a better match in the latter. c) shows a good match to normality for untransformed period estimates from WT plants carrying a CCA1p:LUC (blue) or TOC1p:LUC (yellow) transgene.

Figure 5

Behaviour of the U2019.3 and U2020.3 models, compared to TiMet RNA data for WT plants.

RNA levels for four transcripts in Col-0 WT plants are shown, compared to simulation results from U2019.3 (solid black line) and U2020.3 (dashed pink line). The higher amplitudes of *LHY* and *PRR7* data are better matched by U2020.3. The models were entrained for 10 days in L:D cycles (white:dark grey bars) before transfer to constant light (white:light grey) at time 0. Errors bars for experimental data are 1 S.D., n=2.

Downloaded from https://academic.oup.com/insilicoplants/advance-article/doi/10.1093/insilicoplants/diab022/6342117 by guest on 11 August 2021

Figure 6

Recalibration of the Sidaway-Lee data into units of [transcripts] $[cell]^{-1}$ $[h]^{-1}$.

a) Distribution of expression fold-change values between the two timepoints, for the 96 Arabidopsis transcripts in the Piques data set. The red band highlights the interval of selected transcripts, 1± 0.2. b) Steady state levels of transcripts at 17 ºC in the Sidaway-Lee data. c) Linear regression for the log-transformed levels of selected transcripts from the Piques data against their levels in the Sidaway-Lee data. d) Sidaway-Lee's distribution of transcription rates, transformed using the regression into units of log([transcripts] $[cell]^{-1}$ $[h]^{-1}$), compared to a normal distribution (solid line) with maximum likelihood estimation for the mean (μ=3.65) and standard deviation (σ=1.13) of the transcription rates.

Figure 7

Parameters of RNA metabolism inferred in the models, compared to the observed distributions.

a), b) Maximal transcriptional rate parameters relative to the PETR distribution and c), d) RNA degradation rate constants relative to the rates measured by (Sidaway-Lee *et al.* 2014), for the U2019.3 model (left panels) and U2020.3 model (right panels). cL_mL and cL_mD indicate decay rates for cL_m in light and in darkness respectively. Yellow bars indicate the parameter value for each transcript variable. Solid lines show fitted normal distributions with unitless parameters μ=3.65, σ=1.13 (a, b) and μ=-1.76, σ=0.62 (c, d).

Downloaded from https://academic.oup.com/insilicoplants/advance-article/doi/10.1093/insilicoplants/diab022/6342117 by guest on 11 August 2021

8. Author Contributions

A.J.M. and U.U-G conceived the study and analysed modelling results. U.U-G. performed the model development, derived the PETR distribution, ensured software reproducibility and drafted the manuscript. U.U-G. and A.J.M equally contributed to the final manuscript. The authors declare no conflicts of interest.

9. Acknowledgements

We gratefully acknowledge Anna Flis and Mark Stitt (Max Planck Institute for Molecular Plant Physiology, Potsdam-Golm, Germany) for providing raw qRT-PCR data ahead of publication, and Alastair Hume (EPCC, University of Edinburgh), Ryan Gutenkunst (University of Arizona, USA) and James P. Sethna (Cornell University, USA) for providing software support regarding *ad hoc* model constraints and the parallel processing capability in SloppyCell.

Funding: U.U. was supported by Ph.D. scholarship 216707 from Consejo Nacional de Ciencia y Tecnología (CONACYT, México).

Downloaded from https://academic.oup.com/insilicoplants/advance-article/doi/10.1093/insilicoplants/diab022/6342117 by guest on 11 August 2021

10. References

Adams, R.R., Tsorman, N., Stratford, K., Akman, O.E., Gilmore, S., Juty, N., Le Novere, N., and Millar, A.J. (2012). The Input Signal Step Function (ISSF), a standard method to encode input signals in SBML models with software support, applied to circadian clock models. *J. Biol. Rhythms* 27, 328-332.

Alabadí D, Oyama T, Yanovsky MJ, Harmon FG, Más P, Kay SA. 2001. Reciprocal Regulation Between TOC1 and LHY/CCA1 Within the Arabidopsis Circadian Clock. *Science* 293(5531):880–883.

Andronis C, Barak S, Knowles SM, Sugano S, Tobin EM. 2008. The Clock Protein CCA1 and the bZIP Transcription Factor HY5 Physically Interact to Regulate Gene Expression in Arabidopsis. *Molecular Plant* 1(1):58–67.

Bar-Even A, Noor E, Savir Y, Liebermeister W, Davidi D, Tawfik DS, Milo R. 2011. The Moderately Efficient Enzyme: Evolutionary and Physicochemical Trends Shaping Enzyme Parameters. *Biochemistry* 50(21):4402–4410.

Barnes SL, Belliveau NM, Ireland WT, Kinney JB, Phillips R. 2019. Mapping DNA sequence to transcription factor binding energy in vivo. *PLOS Computational Biology* 15(2):e1006226.

Bendix C, Marshall CM, Harmon FG. 2015. Circadian Clock Genes Universally Control Key Agricultural Traits. *Molecular Plant* 8(8):1135–1152.

Bonnot T, Nagel DH. 2021. Time of day prioritizes the pool of translating mRNAs in response to heat stress. *The Plant Cell*.

Downloaded from https://academic.oup.com/insilicoplants/advance-article/doi/10.1093/insilicoplants/diab022/6342117 by guest on 11 August 2021

Buchler NE, Louis M. 2008. Molecular Titration and Ultrasensitivity in Regulatory Networks. *Journal of Molecular Biology* 384(5):1106–1119.

Bujdoso N, Davis SJ. 2013. Mathematical modeling of an oscillating gene circuit to unravel the circadian clock network of Arabidopsis thaliana. *Frontiers in Plant Science* 4.

Choi K, Medley JK, König M, Stocking K, Smith L, Gu S, Sauro HM. 2018. Tellurium: An extensible python-based modeling environment for systems and synthetic biology. *Biosystems* 171:74–79.

Chow BY, Helfer A, Nusinow DA, Kay SA. 2012. ELF3 recruitment to the PRR9 promoter requires other Evening Complex members in the Arabidopsis circadian clock. *Plant Signaling & Behavior* 7(2):170–173.

Covington MF, Maloof JN, Straume M, Kay SA, Harmer SL. 2008. Global transcriptome analysis reveals circadian regulation of key pathways in plant growth and development. *Genome Biology* 9(8):R130.

Czechowski T, Stitt M, Altmann T, Udvardi MK, Scheible W-R. 2005. Genome-Wide Identification and Testing of Superior Reference Genes for Transcript Normalization in Arabidopsis. *Plant Physiology* 139(1):5–17.

D'Agostino R, Pearson ES. 1973. Tests for Departure from Normality. Empirical Results for the Distributions of b2 and $\sqrt{b1}$. *Biometrika* 60(3):613–622.

D'Agostino RB. 1971. An Omnibus Test of Normality for Moderate and Large Size Samples. *Biometrika* 58(2):341–348.

Davies M. 2001. Knowledge (Explicit and Implicit): Philosophical Aspects. In N. J. Smelser & P. B. Baltes (Eds.), *International Encyclopedia of the Social & Behavioral Sciences* (pp. 8126–8132). Oxford: Pergamon.

Downloaded from https://academic.oup.com/insilicoplants/advance-article/doi/10.1093/insilicoplants/diab022/6342117 by guest on 11 August 2021

De Caluwé J, de Melo JRF, Tosenberger A, Hermans C, Verbruggen N, Leloup J-C, Gonze D. 2017. Modeling the photoperiodic entrainment of the plant circadian clock. *Journal of Theoretical Biology* 420:220–231.

De Caluwé J, Xiao Q, Hermans C, Verbruggen N, Leloup J-C, Gonze D. 2016. A Compact Model for the Complex Plant Circadian Clock. *Frontiers in Plant Science* 7.

De Leone MJ, Hernando CE, Romanowski A, García-Hourquet M, Careno D, Casal J, Rugnone M, Mora-García S, Yanovsky MJ. 2019. The LNK Gene Family: At the Crossroad between Light Signaling and the Circadian Clock. *Genes* 10(1):2.

Edwards KD, Anderson PE, Hall A, Salathia NS, Locke JCW, Lynn JR, Straume M, Smith JQ, Millar AJ. 2006. FLOWERING LOCUS C Mediates Natural Variation in the High-Temperature Response of the Arabidopsis Circadian Clock. *The Plant Cell* 18(3):639–650.

Edwards KD, Takata N, Johansson M, Jurca M, Novák O, Hényková E, Liverani S, Kozarewa I, Strnad M, Millar AJ, Ljung K, Eriksson ME. 2018. Circadian clock components control daily growth activities by modulating cytokinin levels and cell division-associated gene expression in Populus trees. *Plant, Cell & Environment* 41(6):1468–1482.

Ezer D, Jung J-H, Lan H, Biswas S, Gregoire L, Box MS, Charoensawan V, Cortijo S, Lai X, Stöckle D, Zubieta C, Jaeger KE, Wigge PA. 2017. The evening complex coordinates environmental and endogenous signals in *Arabidopsis*. *Nature Plants* 3(7):17087.

Farinas B, Mas P. 2011. Functional implication of the MYB transcription factor RVE8/LCL5 in the circadian control of histone acetylation. *The Plant Journal* 66(2):318–329.

Flis Anna, Fernández Aurora Piñas, Zielinski Tomasz, Mengin Virginie, Sulpice Ronan, Stratford Kevin, Hume Alastair, Pokhilko Alexandra, Southern Megan M., Seaton

Downloaded from https://academic.oup.com/insilicoplants/advance-article/doi/10.1093/insilicoplants/diab022/6342117 by guest on 11 August 2021

Daniel D., McWatters Harriet G., Stitt Mark, Halliday Karen J., Millar Andrew J. 2015. Defining the robust behaviour of the plant clock gene circuit with absolute RNA timeseries and open infrastructure. *Open Biology* 5(10):150042.

Fogelmark K, Troein C. 2014. Rethinking Transcriptional Activation in the Arabidopsis Circadian Clock. *PLOS Computational Biology* 10(7):e1003705.

Foo M, Bates DG, Akman OE. 2020. A simplified modelling framework facilitates more complex representations of plant circadian clocks. *PLOS Computational Biology* 16(3):e1007671.

Foo M, Somers DE, Kim P-J. 2016. Kernel Architecture of the Genetic Circuitry of the Arabidopsis Circadian System. *PLOS Computational Biology* 12(2):e1004748.

Franco-Zorrilla JM, López-Vidriero I, Carrasco JL, Godoy M, Vera P, Solano R. 2014. DNA-binding specificities of plant transcription factors and their potential to define target genes. *Proceedings of the National Academy of Sciences* 111(6):2367–2372.

Fujiwara S, Wang L, Han L, Suh S-S, Salomé PA, McClung CR, Somers DE. 2008. Post-translational Regulation of the Arabidopsis Circadian Clock through Selective Proteolysis and Phosphorylation of Pseudo-response Regulator Proteins. *Journal of Biological Chemistry* 283(34):23073–23083.

Furusawa C, Suzuki T, Kashiwagi A, Yomo T, Kaneko K. 2005. Ubiquity of log-normal distributions in intra-cellular reaction dynamics. *BIOPHYSICS* 1:25–31.

Gendron JM, Pruneda-Paz JL, Doherty CJ, Gross AM, Kang SE, Kay SA. 2012. Arabidopsis circadian clock protein, TOC1, is a DNA-binding transcription factor. *Proceedings of the National Academy of Sciences* 109(8):3167–3172.

Downloaded from https://academic.oup.com/insilicoplants/advance-article/doi/10.1093/insilicoplants/diab022/6342117 by guest on 11 August 2021

Greenwood M, Tokuda IT, Locke JCW. 2020. A spatial model of the plant circadian clock reveals design principles for coordinated timing under noisy environments. *BioRxiv* 2020.09.13.294785.

Harmer SL, Hogenesch JB, Straume M, Chang H-S, Han B, Zhu T, Wang X, Kreps JA, Kay SA. 2000. Orchestrated Transcription of Key Pathways in Arabidopsis by the Circadian Clock. *Science* 290(5499):2110–2113.

Higham CF, Husmeier D. 2013. A Bayesian approach for parameter estimation in the extended clock gene circuit of Arabidopsis thaliana. *BMC Bioinformatics* 14(10):S3.

Hsu PY, Devisetty UK, Harmer SL. 2013. Accurate timekeeping is controlled by a cycling activator in Arabidopsis. *ELife* 2:e00473.

Huang H, Alvarez S, Bindbeutel R, Shen Z, Naldrett MJ, Evans BS, Briggs SP, Hicks LM, Kay SA, Nusinow DA. 2016. Identification of Evening Complex Associated Proteins in Arabidopsis by Affinity Purification and Mass Spectrometry. *Molecular & Cellular Proteomics* 15(1):201–217.

Huang W, Pérez-García P, Pokhilko A, Millar AJ, Antoshechkin I, Riechmann JL, Mas P. 2012. Mapping the Core of the Arabidopsis Circadian Clock Defines the Network Structure of the Oscillator. *Science* 336(6077):75–79.

Hughes ME, Abruzzi KC, Allada R, Anafi R, Arpat AB, Asher G, … Hogenesch JB. 2017. Guidelines for Genome-Scale Analysis of Biological Rhythms. *Journal of Biological Rhythms* 32(5):380–393.

James AB, Syed NH, Bordage S, Marshall J, Nimmo GA, Jenkins GI, Herzyk P, Brown JWS, Nimmo HG. 2012. Alternative Splicing Mediates Responses of the Arabidopsis Circadian Clock to Temperature Changes. *The Plant Cell* 24(3):961–981.

Downloaded from https://academic.oup.com/insilicoplants/advance-article/doi/10.1093/insilicoplants/diab022/6342117 by guest on 11 August 2021

Jolley CC, Ukai-Tadenuma M, Perrin D, Ueda HR. 2014. A Mammalian Circadian Clock Model Incorporating Daytime Expression Elements. *Biophysical Journal* 107(6):1462–1473.

Jorgensen P, Nishikawa JL, Breitkreutz B-J, Tyers M. 2002. Systematic Identification of Pathways That Couple Cell Growth and Division in Yeast. *Science* 297(5580):395–400.

Kamioka M, Takao S, Suzuki T, Taki K, Higashiyama T, Kinoshita T, Nakamichi N. 2016. Direct Repression of Evening Genes by CIRCADIAN CLOCK-ASSOCIATED1 in the Arabidopsis Circadian Clock. *The Plant Cell* 28(3):696–711.

Kang B-C, Yun J-Y, Kim S-T, Shin Y, Ryu J, Choi M, Woo JW, Kim J-S. 2018. Precision genome engineering through adenine base editing in plants. *Nature Plants* 4(7):427–431.

Kim JK, Forger DB. 2012. A mechanism for robust circadian timekeeping via stoichiometric balance. *Molecular Systems Biology* 8(1):630.

Kinney JB, Murugan A, Callan CG, Cox EC. 2010. Using deep sequencing to characterize the biophysical mechanism of a transcriptional regulatory sequence. *Proceedings of the National Academy of Sciences* 107(20):9158–9163.

Krahmer J, Goralogia GS, Kubota A, Zardilis A, Johnson RS, Song YH, MacCoss MJ, Bihan TL, Halliday KJ, Imaizumi T, Millar AJ. 2019. Time-resolved interaction proteomics of the GIGANTEA protein under diurnal cycles in Arabidopsis. *FEBS Letters* 593(3):319–338.

Kreutz C, Rodriguez MMB, Maiwald T, Seidl M, Blum HE, Mohr L, Timmer J. 2007. An error model for protein quantification. *Bioinformatics* 23(20):2747–2753.

Downloaded from https://academic.oup.com/insilicoplants/advance-article/doi/10.1093/insilicoplants/diab022/6342117 by guest on 11 August 2021

Lee T-H, Maheshri N. 2012. A regulatory role for repeated decoy transcription factor binding sites in target gene expression. *Molecular Systems Biology* 8(1):576.

Li M, Cao L, Mwimba M, Zhou Y, Li L, Zhou M, Schnable PS, O'Rourke JA, Dong X, Wang W. 2019. Comprehensive mapping of abiotic stress inputs into the soybean circadian clock. *Proceedings of the National Academy of Sciences* 116(47):23840–23849.

Lin Q, Zong Y, Xue C, Wang S, Jin S, Zhu Z, Wang Y, Anzalone AV, Raguram A, Doman JL, Liu DR, Gao C. 2020. Prime genome editing in rice and wheat. *Nature Biotechnology* 38(5):582–585.

Liu T, Carlsson J, Takeuchi T, Newton L, Farré EM. 2013. Direct regulation of abiotic responses by the Arabidopsis circadian clock component PRR7. *The Plant Journal* 76(1):101–114.

Locke JCW, Millar AJ, Turner MS. 2005a. Modelling genetic networks with noisy and varied experimental data: the circadian clock in Arabidopsis thaliana. *Journal of Theoretical Biology* 234(3):383–393.

Locke JCW, Southern MM, Kozma-Bognár L, Hibberd V, Brown PE, Turner MS, Millar AJ. 2005b. Extension of a genetic network model by iterative experimentation and mathematical analysis. *Molecular Systems Biology* 1(1):2005.0013.

Marshall-Colon A, Long SP, Allen DK, Allen G, Beard DA, Benes B, von Caemmerer S, Christensen AJ, Cox DJ, Hart JC, Hirst PM, Kannan K, Katz DS, Lynch JP, Millar AJ, Panneerselvam B, Price ND, Prusinkiewicz P, Raila D, Shekar RG, Shrivastava S, Shukla D, Srinivasan V, Stitt M, Turk MJ, Voit EO, Wang Y, Yin X, Zhu X-G. 2017. Crops In Silico: Generating Virtual Crops Using an Integrative and Multi-scale Modeling Platform. *Frontiers in Plant Science* 8.

Downloaded from https://academic.oup.com/insilicoplants/advance-article/doi/10.1093/insilicoplants/diab022/6342117 by guest on 11 August 2021

Martelot GL, Canella D, Symul L, Migliavacca E, Gilardi F, Liechti R, Martin O, Harshman K, Delorenzi M, Desvergne B, Herr W, Deplancke B, Schibler U, Rougemont J, Guex N, Hernandez N, Naef F, Consortium the C. 2012. Genome-Wide RNA Polymerase II Profiles and RNA Accumulation Reveal Kinetics of Transcription and Associated Epigenetic Changes During Diurnal Cycles. *PLOS Biology* 10(11):e1001442.

Matsuzaki J, Kawahara Y, Izawa T. 2015. Punctual Transcriptional Regulation by the Rice Circadian Clock under Fluctuating Field Conditions. *The Plant Cell* 27(3):633–648.

Millar AJ. 2016. The Intracellular Dynamics of Circadian Clocks Reach for the Light of Ecology and Evolution. *Annual Review of Plant Biology* 67(1):595–618.

Millar AJ, Urquiza U, Freeman PL, Hume A, Plotkin GD, Sorokina O, Zardilis A, Zielinski T. 2019. Practical steps to digital organism models, from laboratory model species to 'Crops in silico. *Journal of Experimental Botany* 70(9):2403–2418.

Milo R, Jorgensen P, Moran U, Weber G, Springer M. 2010. BioNumbers—the database of key numbers in molecular and cell biology. *Nucleic Acids Research* 38(suppl_1):D750–D753.

Mizoguchi T, Wheatley K, Hanzawa Y, Wright L, Mizoguchi M, Song H-R, Carré IA, Coupland G. 2002. LHY and CCA1 Are Partially Redundant Genes Required to Maintain Circadian Rhythms in Arabidopsis. *Developmental Cell* 2(5):629–641.

Mombaerts L, Mauroy A, Gonçalves J. 2016. Optimising time-series experimental design for modelling of circadian rhythms: the value of transient data. *IFAC-PapersOnLine* 49(26):109–113.

Morohashi M, Winn AE, Borisuk MT, Bolouri H, Doyle J, Kitano H. 2002. Robustness as a Measure of Plausibility in Models of Biochemical Networks. *Journal of Theoretical Biology* 216(1):19–30.

Downloaded from https://academic.oup.com/insilicoplants/advance-article/doi/10.1093/insilicoplants/diab022/6342117 by guest on 11 August 2021

Müller LM, Mombaerts L, Pankin A, Davis SJ, Webb AAR, Goncalves J, Korff M von. 2020. Differential Effects of Day/Night Cues and the Circadian Clock on the Barley Transcriptome. *Plant Physiology* 183(2):765–779.

Müller NA, Wijnen CL, Srinivasan A, Ryngajllo M, Ofner I, Lin T, Ranjan A, West D, Maloof JN, Sinha NR, Huang S, Zamir D, Jiménez-Gómez JM. 2016. Domestication selected for deceleration of the circadian clock in cultivated tomato. *Nature Genetics* 48(1):89–93.

Myers CR, Gutenkunst RN, Sethna JP. 2007. Python Unleashed on Systems Biology. *Computing in Science Engineering* 9(3):34–37.

Nagel DH, Doherty CJ, Pruneda-Paz JL, Schmitz RJ, Ecker JR, Kay SA. 2015. Genome-wide identification of CCA1 targets uncovers an expanded clock network in Arabidopsis. *Proceedings of the National Academy of Sciences* 112(34):E4802–E4810.

Nakamichi N, Kiba T, Henriques R, Mizuno T, Chua N-H, Sakakibara H. 2010. PSEUDO-RESPONSE REGULATORS 9, 7, and 5 Are Transcriptional Repressors in the Arabidopsis Circadian Clock. *The Plant Cell* 22(3):594–605.

Nakamichi N, Kiba T, Kamioka M, Suzuki T, Yamashino T, Higashiyama T, Sakakibara H, Mizuno T. 2012. Transcriptional repressor PRR5 directly regulates clock-output pathways. *Proceedings of the National Academy of Sciences* 109(42):17123–17128.

Nohales MA, Kay SA. 2016. Molecular mechanisms at the core of the plant circadian oscillator. *Nature Structural & Molecular Biology* 23(12):1061–1069.

Nusinow DA, Helfer A, Hamilton EE, King JJ, Imaizumi T, Schultz TF, Farré EM, Kay SA. 2011. The ELF4–ELF3–LUX complex links the circadian clock to diurnal control of hypocotyl growth. *Nature* 475(7356):398–402.

Downloaded from https://academic.oup.com/insilicoplants/advance-article/doi/10.1093/insilicoplants/diab022/6342117 by guest on 11 August 2021

O'Malley RC, Huang SC, Song L, Lewsey MG, Bartlett A, Nery JR, Galli M, Gallavotti A, Ecker JR. 2016. Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape. *Cell* 165(5):1280–1292.

Pelechano V, Chávez S, Pérez-Ortín JE. 2010. A Complete Set of Nascent Transcription Rates for Yeast Genes. *PLOS ONE* 5(11):e15442.

Pérez-García P, Ma Y, Yanovsky MJ, Mas P. 2015. Time-dependent sequestration of RVE8 by LNK proteins shapes the diurnal oscillation of anthocyanin biosynthesis. *Proceedings of the National Academy of Sciences* 112(16):5249–5253.

Piques M, Schulze WX, Höhne M, Usadel B, Gibon Y, Rohwer J, Stitt M. 2009. Ribosome and transcript copy numbers, polysome occupancy and enzyme dynamics in Arabidopsis. *Molecular Systems Biology* 5(1):314.

Pokhilko A, Fernández AP, Edwards KD, Southern MM, Halliday KJ, Millar AJ. 2012. The clock gene circuit in Arabidopsis includes a repressilator with additional feedback loops. *Molecular Systems Biology* 8(1):574.

Pokhilko A, Hodge SK, Stratford K, Knox K, Edwards KD, Thomson AW, Mizuno T, Millar AJ. 2010. Data assimilation constrains new connections and components in a complex, eukaryotic circadian clock model. *Molecular Systems Biology* 6(1):416.

Pokhilko A, Mas P, Millar AJ. 2013. Modelling the widespread effects of TOC1 signalling on the plant circadian clock and its outputs. *BMC Systems Biology* 7(1):23.

Preuss SB, Meister R, Xu Q, Urwin CP, Tripodi FA, Screen SE, Anil VS, Zhu S, Morrell JA, Liu G, Ratcliffe OJ, Reuber TL, Khanna R, Goldman BS, Bell E, Ziegler TE, McClerren AL, Ruff TG, Petracek ME. 2012. Expression of the Arabidopsis thaliana BBX32 Gene in Soybean Increases Grain Yield. *PLOS ONE* 7(2):e30717.

Downloaded from https://academic.oup.com/insilicoplants/advance-article/doi/10.1093/insilicoplants/diab022/6342117 by guest on 11 August 2021

Raue A, Schilling M, Bachmann J, Matteson A, Schelke M, Kaschek D, Hug S, Kreutz C, Harms BD, Theis FJ, Klingmüller U, Timmer J. 2013. Lessons Learned from Quantitative Dynamical Modeling in Systems Biology. *PLOS ONE* 8(9):e74335.

Rawat R, Takahashi N, Hsu PY, Jones MA, Schwartz J, Salemi MR, Phinney BS, Harmer SL. 2011. REVEILLE8 and PSEUDO-REPONSE REGULATOR5 Form a Negative Feedback Loop within the Arabidopsis Circadian Clock. *PLOS Genetics* 7(3):e1001350.

Rugnone ML, Soverna AF, Sanchez SE, Schlaen RG, Hernando CE, Seymour DK, Mancini E, Chernomoretz A, Weigel D, Más P, Yanovsky MJ. 2013. LNK genes integrate light and clock signaling networks at the core of the Arabidopsis oscillator. *Proceedings of the National Academy of Sciences* 110(29):12120–12125.

Sanchez SE, Rugnone ML, Kay SA. 2020. Light Perception: A Matter of Time. *Molecular Plant* 13(3):363–385.

Schaffer R, Ramsay N, Samach A, Corden S, Putterill J, Carré IA, Coupland G. 1998. The late elongated hypocotyl Mutation of Arabidopsis Disrupts Circadian Rhythms and the Photoperiodic Control of Flowering. *Cell* 93(7):1219–1229.

Seaton DD, Toledo-Ortiz G, Ganpudi A, Kubota A, Imaizumi T, Halliday KJ. 2018. Dawn and photoperiod sensing by phytochrome A. *Proceedings of the National Academy of Sciences* 115(41):10523–10528.

Shalit-Kaneh A, Kumimoto RW, Filkov V, Harmer SL. 2018. Multiple feedback loops of the Arabidopsis circadian clock provide rhythmic robustness across environmental conditions. *Proceedings of the National Academy of Sciences* 115(27):7147–7152.

Downloaded from https://academic.oup.com/insilicoplants/advance-article/doi/10.1093/insilicoplants/diab022/6342117 by guest on 11 August 2021

Sidaway-Lee K, Costa MJ, Rand DA, Finkenstadt B, Penfield S. 2014. Direct measurement of transcription rates reveals multiple mechanisms for configuration of the Arabidopsisambient temperature response. *Genome Biology* 15(3):R45.

Somers DE. 2012. The Arabidopsis clock: time for an about-face? *Genome Biology* 13(4):153.

Urquiza-García U, Millar AJ. 2019. Expanding the bioluminescent reporter toolkit for plant science with NanoLUC. *Plant Methods* 15(1):68.

Wang Z-Y, Tobin EM. 1998. Constitutive Expression of the CIRCADIAN CLOCK ASSOCIATED 1 (CCA1) Gene Disrupts Circadian Rhythms and Suppresses Its Own Expression. *Cell* 93(7):1207–1217.

Weirauch MT, Yang A, Albu M, Cote AG, Montenegro-Montero A, Drewe P, Najafabadi HS, Lambert SA, Mann I, Cook K, Zheng H, Goity A, van Bakel H, Lozano J-C, Galli M, Lewsey MG, Huang E, Mukherjee T, Chen X, Reece-Hoyes JS, Govindarajan S, Shaulsky G, Walhout AJM, Bouget F-Y, Ratsch G, Larrondo LF, Ecker JR, Hughes TR. 2014. Determination and Inference of Eukaryotic Transcription Factor Sequence Specificity. *Cell* 158(6):1431–1443.

Wuyts N, Palauqui J-C, Conejero G, Verdeil J-L, Granier C, Massonnet C. 2010. High-contrast three-dimensional imaging of the Arabidopsis leaf enables the analysis of cell dimensions in the epidermis and mesophyll. *Plant Methods* 6(1):17.

Xie Q, Wang P, Liu X, Yuan L, Wang L, Zhang C, Li Y, Xing H, Zhi L, Yue Z, Zhao C, McClung CR, Xu X. 2014. LNK1 and LNK2 Are Transcriptional Coactivators in the Arabidopsis Circadian Oscillator. *The Plant Cell* 26(7):2843–2857.

Xu R, Li J, Liu X, Shan T, Qin R, Wei P. 2020. Development of Plant Prime-Editing Systems for Precise Genome Editing. *Plant Communications* 1(3):100043.

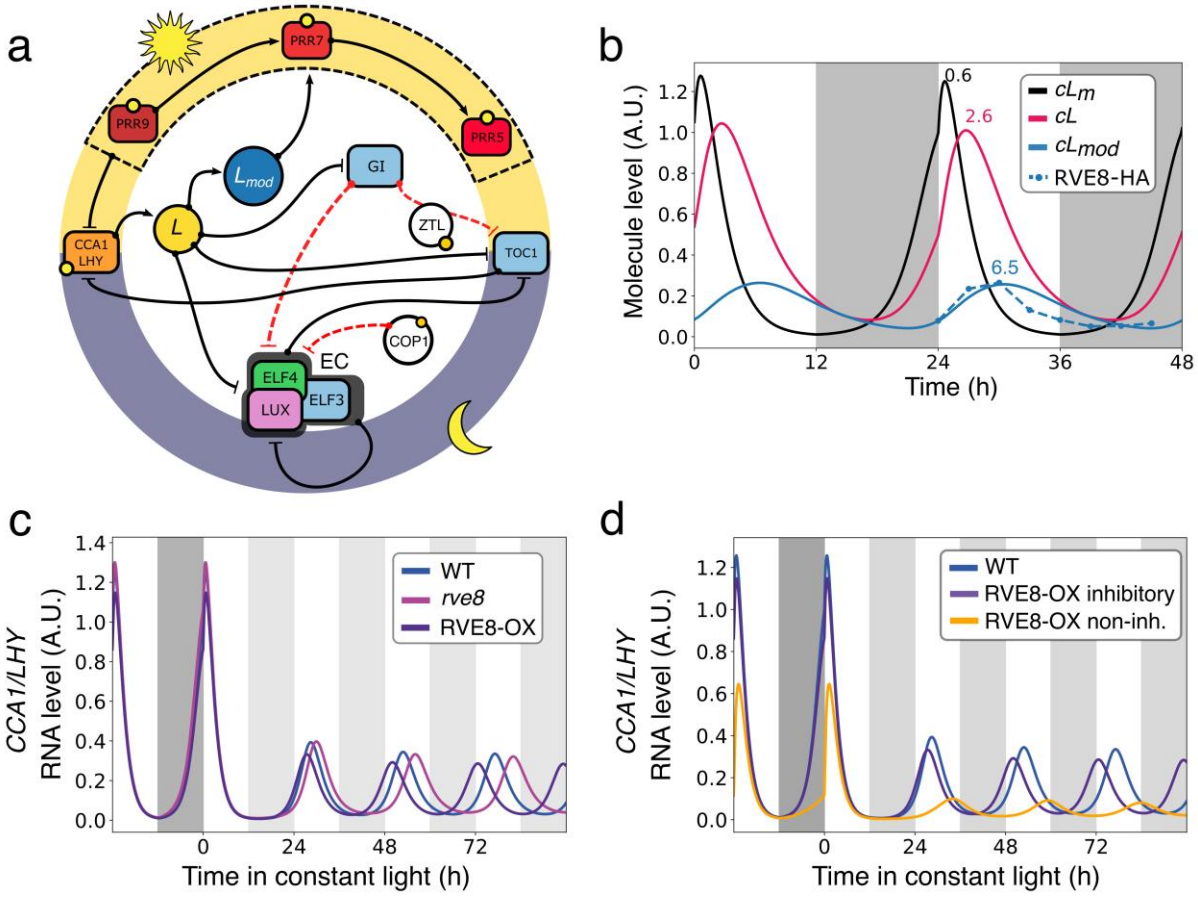
Downloaded from https://academic.oup.com/insilicoplants/advance-article/doi/10.1093/insilicoplants/diab022/6342117 by guest on 11 August 2021

Yakir E, Hilman D, Hassidim M, Green RM. 2007. CIRCADIAN CLOCK ASSOCIATED1 Transcript Stability and the Entrainment of the Circadian Clock in Arabidopsis. *Plant Physiology* 145(3):925–932.

Zeilinger MN, Farré EM, Taylor SR, Kay SA, Doyle III FJ. 2006. A novel computational model of the circadian clock in Arabidopsis that incorporates PRR7 and PRR9. *Molecular Systems Biology* 2(1):58.

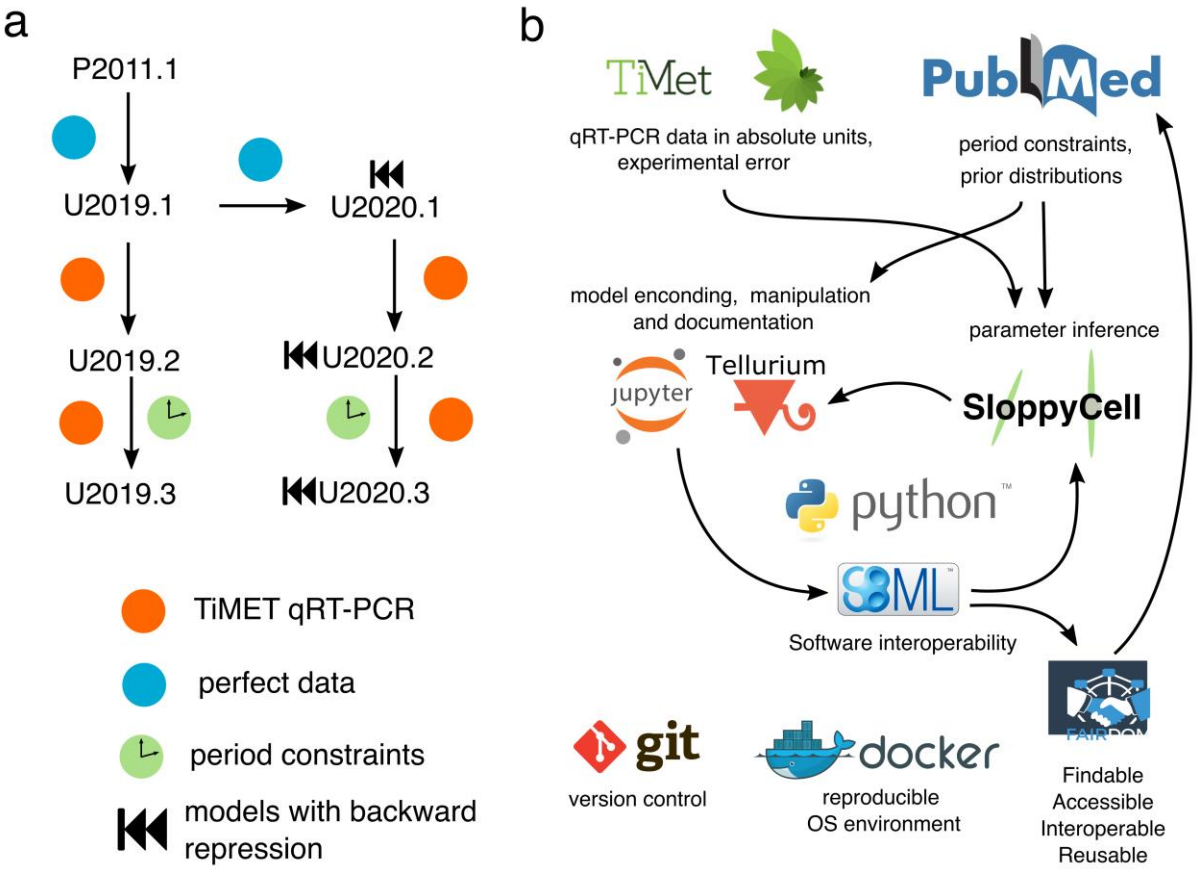
Downloaded from https://academic.oup.com/insilicoplants/advance-article/doi/10.1093/insilicoplants/diab022/6342117 by guest on 11 August 2021

Figure 1



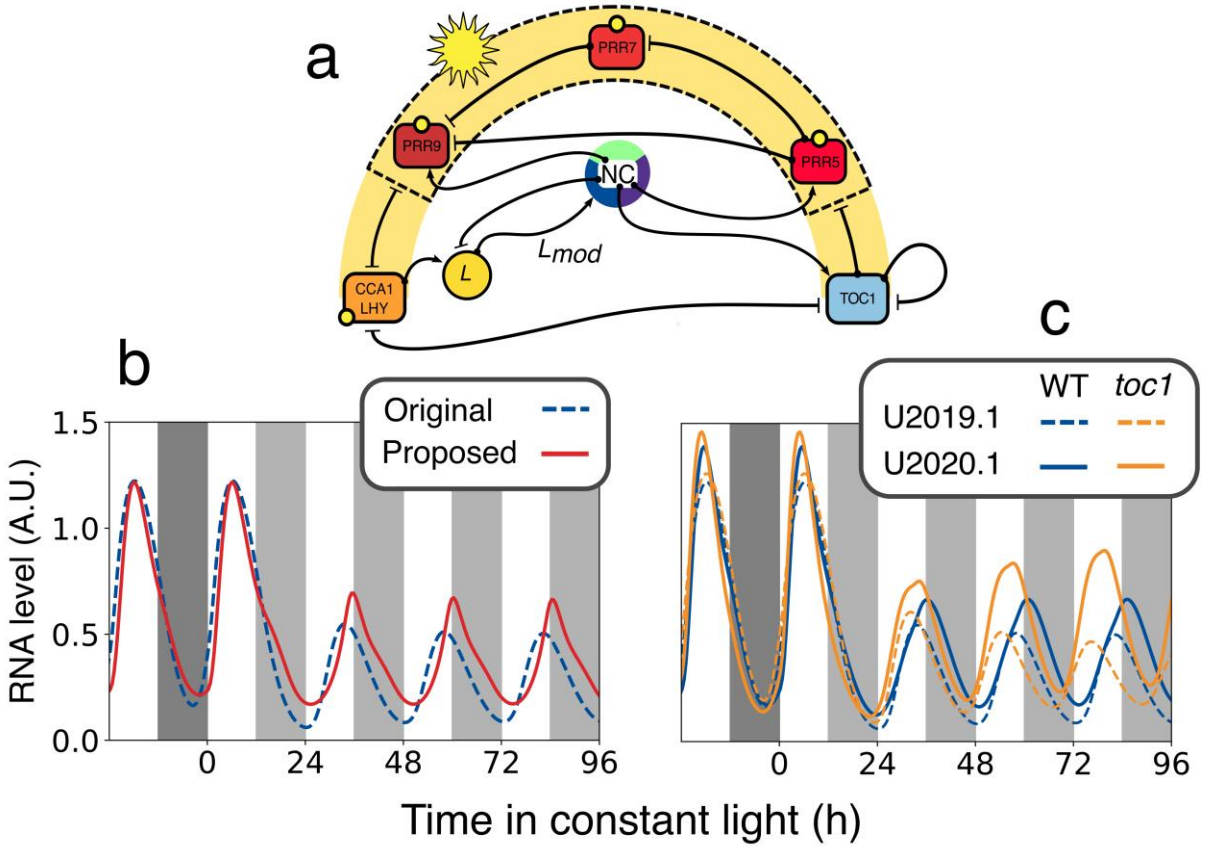
Downloaded from https://academic.oup.com/insilicoplants/advance-article/doi/10.1093/insilicoplants/diab022/6342117 by guest on 11 August 2021

Figure 2



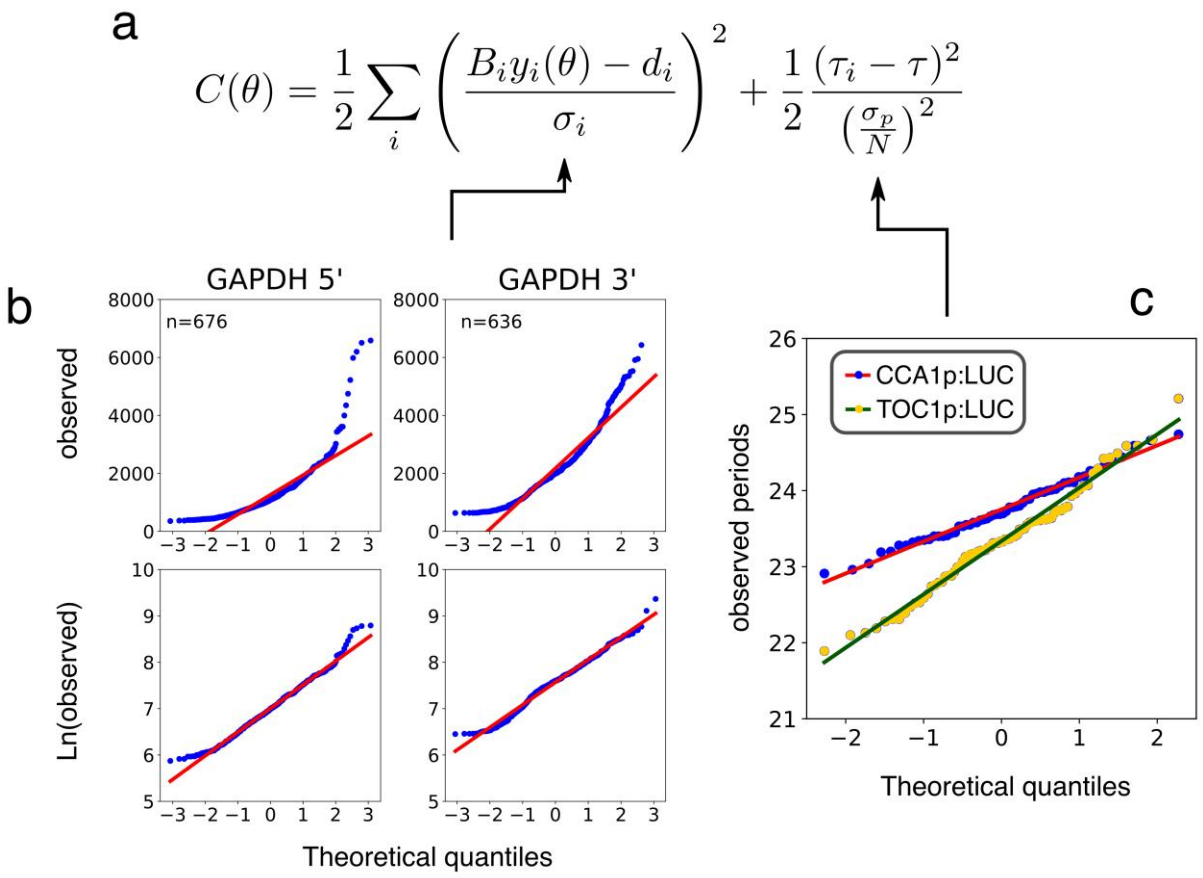
Downloaded from https://academic.oup.com/insilicoplants/advance-article/doi/10.1093/insilicoplants/diab022/6342117 by guest on 11 August 2021

Figure 3



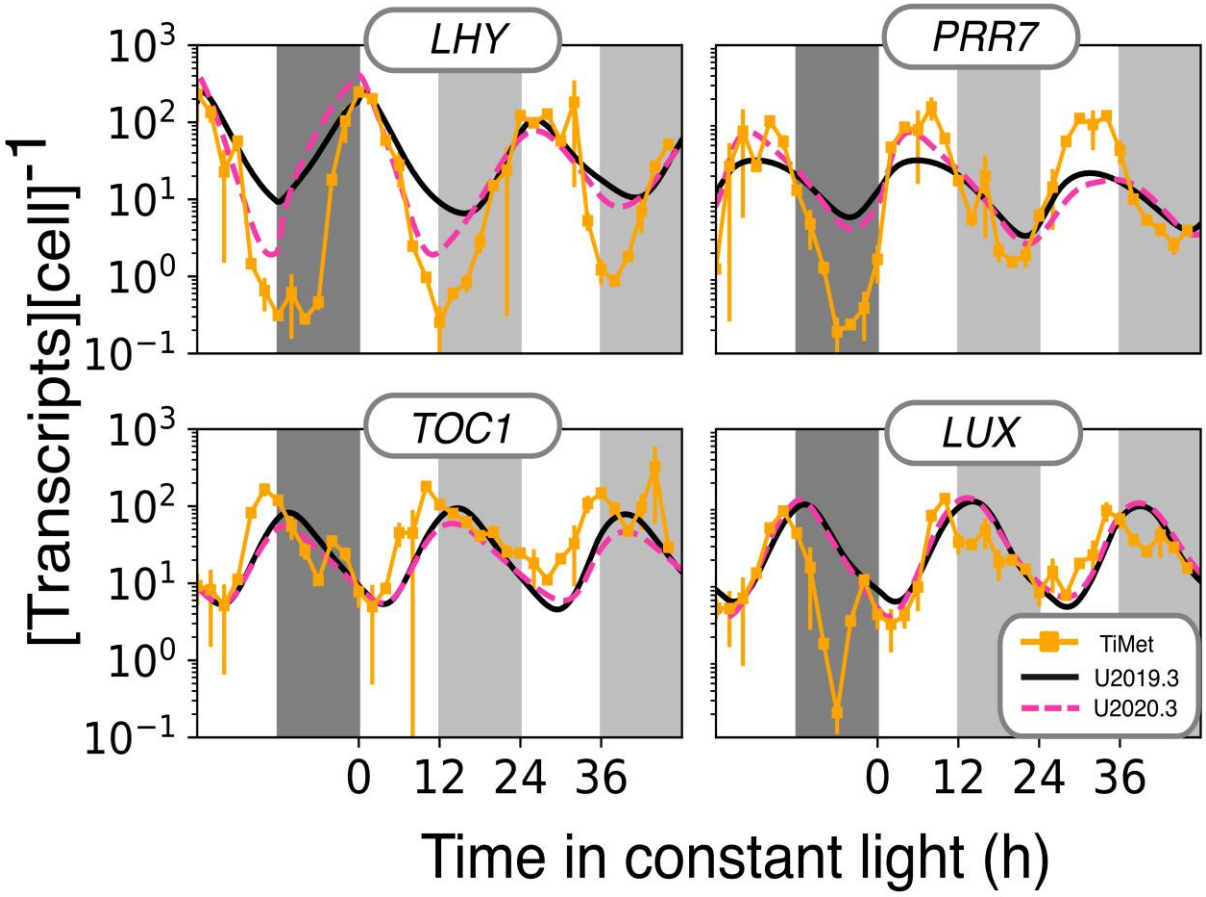
Downloaded from https://academic.oup.com/insilicoplants/advance-article/doi/10.1093/insilicoplants/diab022/6342117 by guest on 11 August 2021

Figure 4



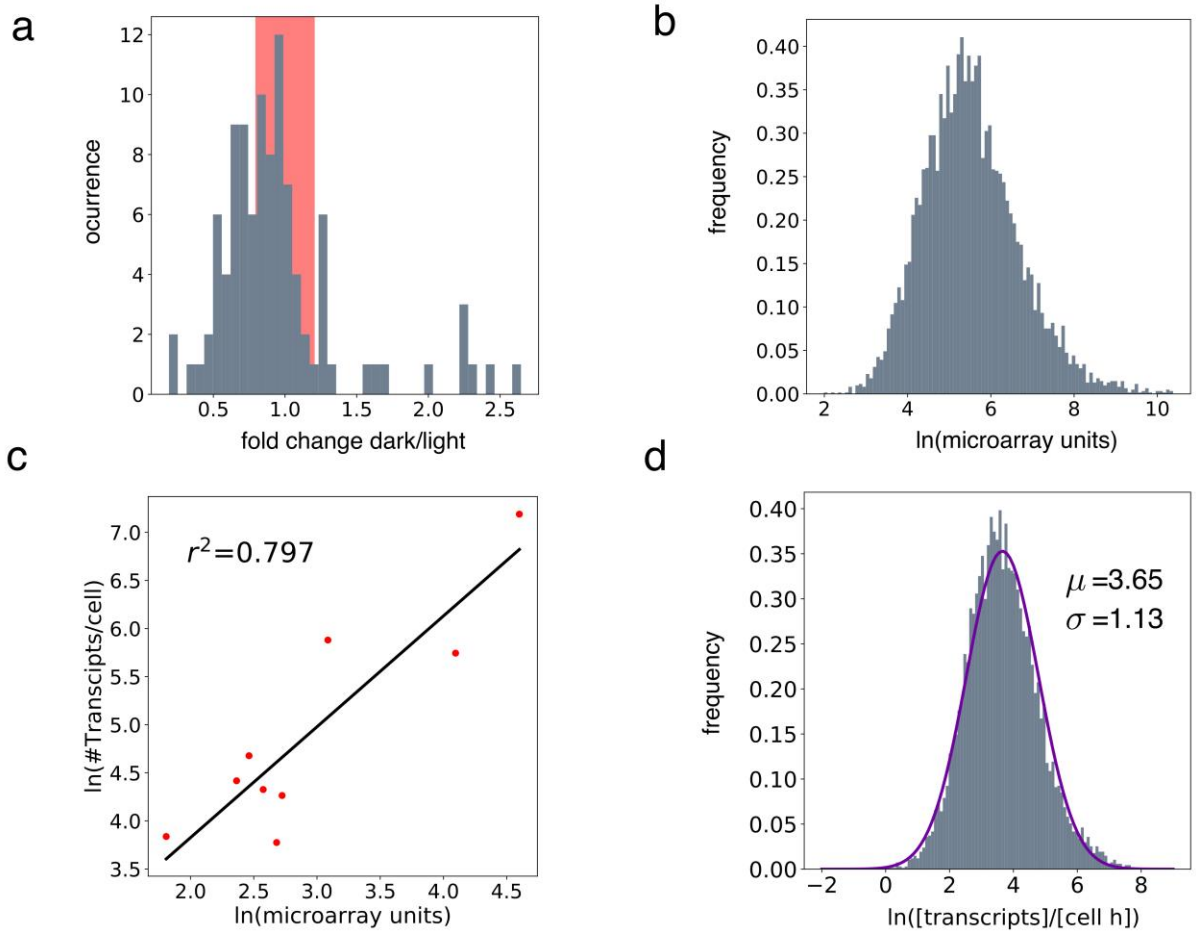
Downloaded from https://academic.oup.com/insilicoplants/advance-article/doi/10.1093/insilicoplants/diab022/6342117 by guest on 11 August 2021

Figure 5



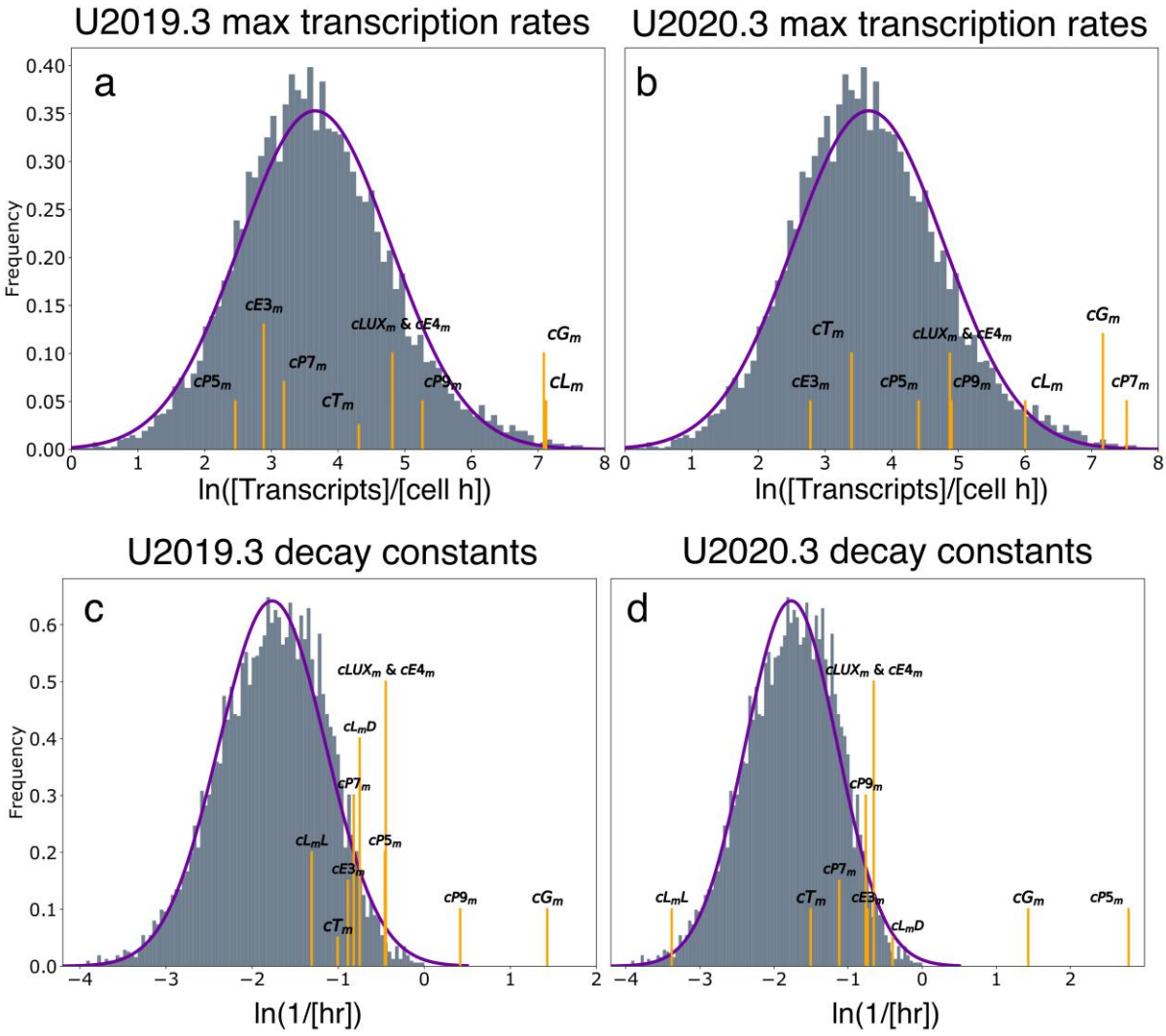
Downloaded from https://academic.oup.com/insilicoplants/advance-article/doi/10.1093/insilicoplants/diab022/6342117 by guest on 11 August 2021

Figure 6



Downloaded from https://academic.oup.com/insilicoplants/advance-article/doi/10.1093/insilicoplants/diab022/6342117 by guest on 11 August 2021

Figure 7



Downloaded from https://academic.oup.com/insilicoplants/advance-article/doi/10.1093/insilicoplants/diab022/6342117 by guest on 11 August 2021