Supplementary Material

Contents

S1 Fitting logic models to data - further details

S1.1 Specification of the cost function

The cost score for a given combination of signalling delays $\tau = (\tau_1, \ldots, \tau_{N+m})$, thresholds $\mathbf{T} = (T_1, \ldots, T_n)$ and logic gates $\mathbf{G} = (g_1, \ldots, g_d)$ is calculated from the target data set as follows:

Let D be an $n \times q$ array comprising the original data and $\hat{D}_{\mathbf{T}}$ the corresponding discretised expression data obtained using the thresholds $\mathbf{T}$. Analogously, we introduce R as an $n \times r$ array of continuous random numbers in the interval $[-L, L]$, where L is arbitrary and $r \gg q$. R is discretised using the same choice of thresholds $\mathbf{T}$ to obtain an array of discrete random expression profiles $\hat{R}_{\mathbf{T}}$.

The correlation score, c_i, for the ith vertex is defined to be the difference between the correlations obtained using $\hat{D}_{\mathbf{T}}$ and $\hat{R}_{\mathbf{T}}$:

$$c_i\left(\hat{D}_{\mathbf{T}}, \hat{R}_{\mathbf{T}}, \tau, \mathbf{G}\right) = \tilde{c}_i\left(\hat{D}_{\mathbf{T}}, \tau, \mathbf{G}\right) - \tilde{c}_i\left(\hat{R}_{\mathbf{T}}, \tau, \mathbf{G}\right). \tag{S1}$$

In this expression $\tilde{c}_i\left(\hat{X}, \tau, \mathbf{G}\right)$ is a function that predicts the output of the model at i using the data $\hat{X}$, and then calculates the normalised dot product of this with the data at i. Subtracting the correlation score obtained for the random data set in eqn. (S1) normalises c_i against spurious correlations introduced by the discretisation process.

To compute $\tilde{c}_i$, we define $\hat{X}_i$ to be the vector of discrete expression data for the ith component, and the predicted output $\hat{P}_i$ to be the vector obtained by applying the corresponding Boolean function s_i to $\hat{X}$. As predictions can only be made for $t \geqslant \hat{\tau}_i$, where $\hat{\tau}_i$ is the maximum signalling delay that contributes to s_i, the elements of $\hat{P}_i$ and $\hat{X}_i$ corresponding to $t < \hat{\tau}_i$ are flagged as unknowns. We do this by setting the values arbitrarily to 2.

Removing the elements of $\hat{P}_i$ and $\hat{X}_i$ for which $\hat{P}_i = 2$ and mapping the truncated vectors from $\{0, 1\}$ to $\{-1, 1\}$, via $1 \to 1$ and $0 \to -1$, we denote the resulting vectors $\bar{P}_i$ and $\bar{X}_i$. We calculate

the correlation as

$$\tilde{c}_i\left(\hat{X},\tau,\mathbf{G}\right)=\frac{1}{\left|\bar{X}_i\right|}\bar{X}_i\cdot\bar{P}_i, \tag{S2}$$

where $\cdot$ denotes dot product and $\left|\bar{X}_i\right|$ the cardinality of $\bar{X}_i$. By construction, $\tilde{c}_i$ lies in the interval $[-1,1]$ where 1, 0 and -1 denote correlation, no correlation and anticorrelation, respectively.

The cost score $C\left(\tau,\mathbf{T},\mathbf{G}\right)$ for the data set $\hat{D}_{\mathbf{T}}$ is then computed from (S1) and (S2) as

$$C\left(\tau,\mathbf{T},\mathbf{G}\right)=n-\sum_{i=1}^{n}c_i\left(\hat{D}_{\mathbf{T}},\hat{R}_{\mathbf{T}},\tau,\mathbf{G}\right). \tag{S3}$$

An optimal score of 0 corresponds to a combination of parameters and gates $(\tau,\mathbf{T},\mathbf{G})$ that predicts the time courses exactly.

For the fits of the models to synthetic data, $C\left(\tau,\mathbf{T},\mathbf{G}\right)$ was calculated for both the LD data set, D^{LD}, and the appropriate free-running data set, D^{DD} (*Neurospora*) or D^{LL} (*Arabidopsis*). The overall cost in each case was taken to be the linear sum of these values:

$$C_{\mathrm{Nc}}\left(\tau,\mathbf{T},\mathbf{G}\right) = 2n-\sum_{i=1}^{n}c_i\left(\hat{D}^{DD}_{\mathbf{T}},\hat{R}^{DD}_{\mathbf{T}},\tau,\mathbf{G}\right)-\sum_{i=1}^{n}c_i\left(\hat{D}^{LD}_{\mathbf{T}},\hat{R}^{LD}_{\mathbf{T}},\tau,\mathbf{G}\right), \tag{S4}$$

$$C_{\mathrm{At}}\left(\tau,\mathbf{T},\mathbf{G}\right) = 2n-\sum_{i=1}^{n}c_i\left(\hat{D}^{LL}_{\mathbf{T}},\hat{R}^{LL}_{\mathbf{T}},\tau,\mathbf{G}\right)-\sum_{i=1}^{n}c_i\left(\hat{D}^{LD}_{\mathbf{T}},\hat{R}^{LD}_{\mathbf{T}},\tau,\mathbf{G}\right). \tag{S5}$$

The cost function used for the optimisation of the 3-loop *Arabidopsis* model to LL experimental LUC data, D^{LUC}, was as follows:

$$C^{LUC}_{\mathrm{At}}\left(\tau,\mathbf{T},\mathbf{G}\right)=n-\sum_{i=1}^{n}c_i\left(\hat{D}^{LUC}_{\mathbf{T}},\hat{R}^{LUC}_{\mathbf{T}},\tau,\mathbf{G}\right). \tag{S6}$$

Finally, in order to facilitate comparison across the different models, the scores plotted in Figs. 4, 5 and 8 are linearly scaled so that 0 and 1 represent the lowest and highest optimal values across all the gates considered in each case.

S1.2 Degeneracy of the cost function

Depending on the model and data set, there can be degeneracies in the cost function that make different LCs difficult to distinguish. To illustrate this, we consider the abstract topology of the 1-loop *Neurospora* model shown in Fig. 2A. With discrete time courses of period Q that have equal durations of 0 and 1 states, two different LC-delay combinations will give the same cost score if they differ by the substitution of activation with inhibition (or *vice versa*) in one interaction, and the substitution $\tau\rightarrow\mathrm{mod}(\tau+\frac{Q}{2},Q)$ in the corresponding signalling delay. The 1-loop *Neurospora* model has 4 LCs associated with its adjacency matrix and so a search with no constraints using data from just one light condition will yield a 4-fold degeneracy in the cost function. This can be seen in Supp. Fig. S6A.

For the same model, if we consider discretisations $\mathbf{T}$ that yield discrete time courses with heavily asymmetric durations of 1 and 0 states, we see that the degeneracy partially lifts. In this case, two different parameter combinations will yield similar costs if they are related by the substitution of activation with inhibition (or *vice versa*) in one interaction, the substitution $\tau\rightarrow\mathrm{mod}(\tau+\frac{Q}{2},Q)$ and a change to the discretisation threshold that switches the duration of the 1 and 0 states. However the change of discretisation threshold has consequences for the other interactions in a closed loop, requiring similar changes to be made to the local parameters there too. As a result,

for this discretisation in 1-loop *Neurospora*, the degeneracy in the cost function lifts to a 2-fold degeneracy, as shown in Supp. Fig. S6B.

Experimental data generally incorporates a degree of variability due to the inherent stochasticity of the interactions. Depending on the degree of variability, this introduces minor variations in the expression levels that correlate between upstream and downstream components. These micro-correlations help to lift the degeneracies in the cost function, and data with a greater degree of stochasticity cause a greater lift in the degeneracy. This results in distinct ranked cost profiles for different LCs (Supp. Figs. S6C-H). Similar results were obtained with the other models considered in this study (data not shown).

Taken together, these results indicate that cost function degeneracies can be lifted by increasing the total number of bitstrings costed, either by varying the discretisation thresholds or introducing stochasticity into the data set. This suggests that in order to distinguish different LCs from the experimental and synthetic data sets, both of which exhibit stochasticity, it is necessary to employ an optimisation strategy that is capable of searching over a sufficiently large set of thresholds.

S2 Logic models of the clock networks

This section presents the logic versions of the clock models used in this study. In each case, both the general and optimal logic models are given (the latter correspond to the logic gate configurations yielding the best fits to data presented in Tables 1 and 2). The following logical symbols are employed for brevity: $\overline{Z}$ = NOT Z (logical complement); $Z_1 \cdot Z_2 = Z_1$AND Z_2 (logical product); $Z_1 + Z_2 = Z_1$OR Z_2 (logical sum) [1, 2].

S2.1 General logic circuits

1-loop *Neurospora*

$$\begin{aligned} FRQ\left(t\right) &= G_1\left(\mathrm{FRQ}\left(t-\tau_2\right), g_2\right) + L_1\left(t-\tau_3\right) \\ \mathrm{FRQ}\left(t\right) &= G_1\left(FRQ\left(t-\tau_1\right), g_1\right) \end{aligned}$$

2-loop *Neurospora*

$$\begin{aligned} FRQ\left(t\right) &= G_2\left(G_1\left(\mathrm{FRQ}_1\left(t-\tau_3\right), g_3\right), G_1\left(\mathrm{FRQ}_2\left(t-\tau_4\right), g_4\right), g_5\right) + L_1\left(t-\tau_5\right) \\ \mathrm{FRQ}_1\left(t\right) &= G_1\left(FRQ\left(t-\tau_1\right), g_1\right) \\ \mathrm{FRQ}_2\left(t\right) &= G_1\left(FRQ\left(t-\tau_2\right), g_2\right) \end{aligned}$$

2-loop *Arabidopsis*

$$\begin{aligned} LHY\left(t\right) &= G_1\left(X\left(t-\tau_3\right), g_3\right) \cdot L_1\left(t-\tau_7\right) \\ TOC1\left(t\right) &= G_2\left(G_1\left(LHY\left(t-\tau_1\right), g_1\right), G_1\left(Y\left(t-\tau_6\right), g_6\right), g_8\right) \\ X\left(t\right) &= G_1\left(TOC1\left(t-\tau_2\right), g_2\right) \\ Y\left(t\right) &= G_2\left(G_1\left(LHY\left(t-\tau_4\right), g_4\right), G_1\left(TOC1\left(t-\tau_5\right), g_5\right), g_7\right) \\ &\qquad \cdot \left(L_2\left(t-\tau_8\right) + L_3\left(t-\tau_9\right)\right) \end{aligned}$$

3-loop *Arabidopsis*

$$
\begin{aligned}
LHY\left(t\right) &= G_2\left(G_1\left(X\left(t-\tau_3\right),g_3\right),G_1\left(PRR\left(t-\tau_8\right),g_{10}\right),g_{11}\right)\cdot L_1\left(t-\tau_9\right)\\
TOC1\left(t\right) &= G_2\left(G_1\left(LHY\left(t-\tau_1\right),g_1\right),G_1\left(Y\left(t-\tau_6\right),g_6\right),g_8\right)\\
X\left(t\right) &= G_1\left(TOC1\left(t-\tau_2\right),g_2\right)\\
Y\left(t\right) &= G_2\left(G_1\left(LHY\left(t-\tau_4\right),g_4\right),G_1\left(TOC1\left(t-\tau_5\right),g_5\right),g_7\right)\\
&\qquad\qquad\cdot\left(L_2\left(t-\tau_{10}\right)+L_3\left(t-\tau_{11}\right)\right)\\
PRR\left(t\right) &= G_1\left(LHY\left(t-\tau_7\right),g_9\right)\cdot L_4\left(t-\tau_{12}\right)
\end{aligned}
$$

S2.2 Optimal networks for synthetic data

1-loop *Neurospora*

$$
\begin{aligned}
FRQ\left(t\right) &= \overline{\mathrm{FRQ}}\left(t-\tau_2\right)+L_1\left(t-\tau_3\right)\\
\mathrm{FRQ}\left(t\right) &= FRQ\left(t-\tau_1\right)
\end{aligned}
$$

2-loop *Neurospora*

$$
\begin{aligned}
FRQ\left(t\right) &= \left(\overline{\mathrm{FRQ}}_1\left(t-\tau_3\right)\cdot\overline{\mathrm{FRQ}}_2\left(t-\tau_4\right)\right)+L_1\left(t-\tau_5\right)\\
\mathrm{FRQ}_1\left(t\right) &= FRQ\left(t-\tau_1\right)\\
\mathrm{FRQ}_2\left(t\right) &= FRQ\left(t-\tau_2\right)
\end{aligned}
$$

2-loop *Arabidopsis*

$$
\begin{aligned}
LHY\left(t\right) &= X\left(t-\tau_3\right)\cdot L_1\left(t-\tau_7\right)\\
TOC1\left(t\right) &= \overline{LHY}\left(t-\tau_1\right)\cdot Y\left(t-\tau_6\right)\\
X\left(t\right) &= TOC1\left(t-\tau_2\right)\\
Y\left(t\right) &= \left(\overline{LHY}\left(t-\tau_4\right)\cdot\overline{TOC1}\left(t-\tau_5\right)\right)\cdot\left(L_2\left(t-\tau_8\right)+L_3\left(t-\tau_9\right)\right)
\end{aligned}
$$

3-loop *Arabidopsis*

$$
\begin{aligned}
LHY\left(t\right) &= \left(X\left(t-\tau_3\right)\cdot\overline{PRR}\left(t-\tau_8\right)\right)\cdot L_1\left(t-\tau_9\right)\\
TOC1\left(t\right) &= \overline{LHY}\left(t-\tau_1\right)\cdot Y\left(t-\tau_6\right)\\
X\left(t\right) &= TOC1\left(t-\tau_2\right)\\
Y\left(t\right) &= \left(\overline{LHY}\left(t-\tau_4\right)\cdot\overline{TOC1}\left(t-\tau_5\right)\right)\cdot\left(L_2\left(t-\tau_{10}\right)+L_3\left(t-\tau_{11}\right)\right)\\
PRR\left(t\right) &= LHY\left(t-\tau_7\right)\cdot L_4\left(t-\tau_{12}\right)
\end{aligned}
$$

S2.3 Optimal networks for experimental *LUC* data

Note that the equations for 3-loop *Arabidopsis* below model the free-running (LL) circuit. Consequently, light inputs and their associated parameters have been removed for brevity. The corresponding circuit diagrams are plotted in Supp. Fig. S4.

Highest ranked LC, G_{OPT}

$$\begin{aligned} LHY\,(t) &= \overline{X}\,(t-\tau_3)\cdot\overline{PRR}\,(t-\tau_8) \\ TOC1\,(t) &= \overline{LHY}\,(t-\tau_1)\cdot Y\,(t-\tau_6) \\ X\,(t) &= TOC1\,(t-\tau_2) \\ Y\,(t) &= LHY\,(t-\tau_4)\cdot\overline{TOC1}(t-\tau_5) \\ PRR\,(t) &= LHY\,(t-\tau_7) \end{aligned}$$

Second-highest ranked LC, G_{DE}

$$\begin{aligned} LHY\,(t) &= X\,(t-\tau_3)\cdot\overline{PRR}\,(t-\tau_8) \\ TOC1\,(t) &= \overline{LHY}\,(t-\tau_1)\cdot Y\,(t-\tau_6) \\ X\,(t) &= TOC1\,(t-\tau_2) \\ Y\,(t) &= \overline{LHY}\,(t-\tau_4)\cdot\overline{TOC1}(t-\tau_5) \\ PRR\,(t) &= LHY\,(t-\tau_7) \end{aligned}$$

References

[1] Thomas, R. 1991 Regulatory networks seen as asynchronous automata: A logical description. *J. Theor. Biol.* **153**, 1–23.

[2] Kaufman, M., Andris, F., Leo, O. 1999 A logical analysis of T cell activation and anergy. *Proc. Natl. Acad. Sci. USA* **7**, 3894–3899.

Supplementary Tables

	DE model	Logic model
One-loop *Neurospora*	13	5
Two-loop *Neurospora*	18	8
Two-loop *Arabidopsis*	64	15
Three-loop *Arabidopsis*	80	20

Table S1: The number of parameters in the differential equation (DE) and logic formulations of each clock circuit. For the logic formulations, this is calculated as the sum of the numbers of delays and discretisation thresholds, together with the number of light inputs possessing a variable pulse length (i.e. for which $p_k \neq P$).

Supplementary Figures

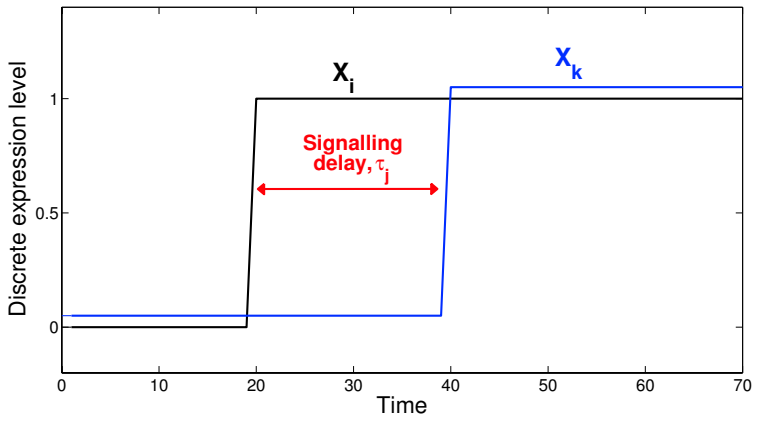

Figure S1: The discrete time course of an upstream entity X_i (black) which activates a downstream entity X_k (blue). The signalling delay τ_j is the time difference (red).

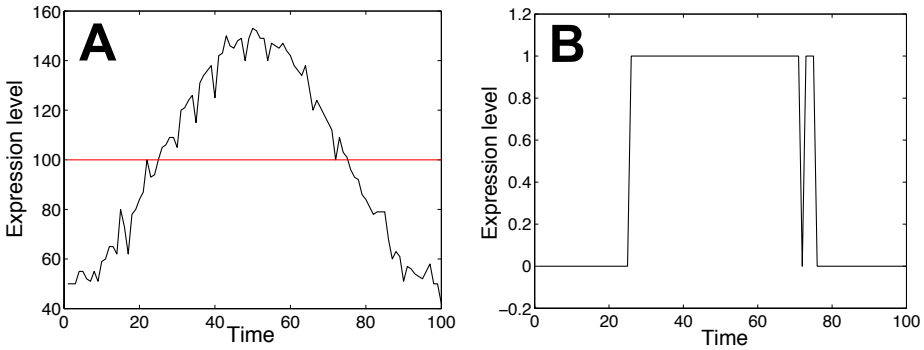

Figure S2: **A**. Discretisation of a hypothetical expression level time course (black) using the threshold shown in red. **B**. The resulting discrete expression time course.

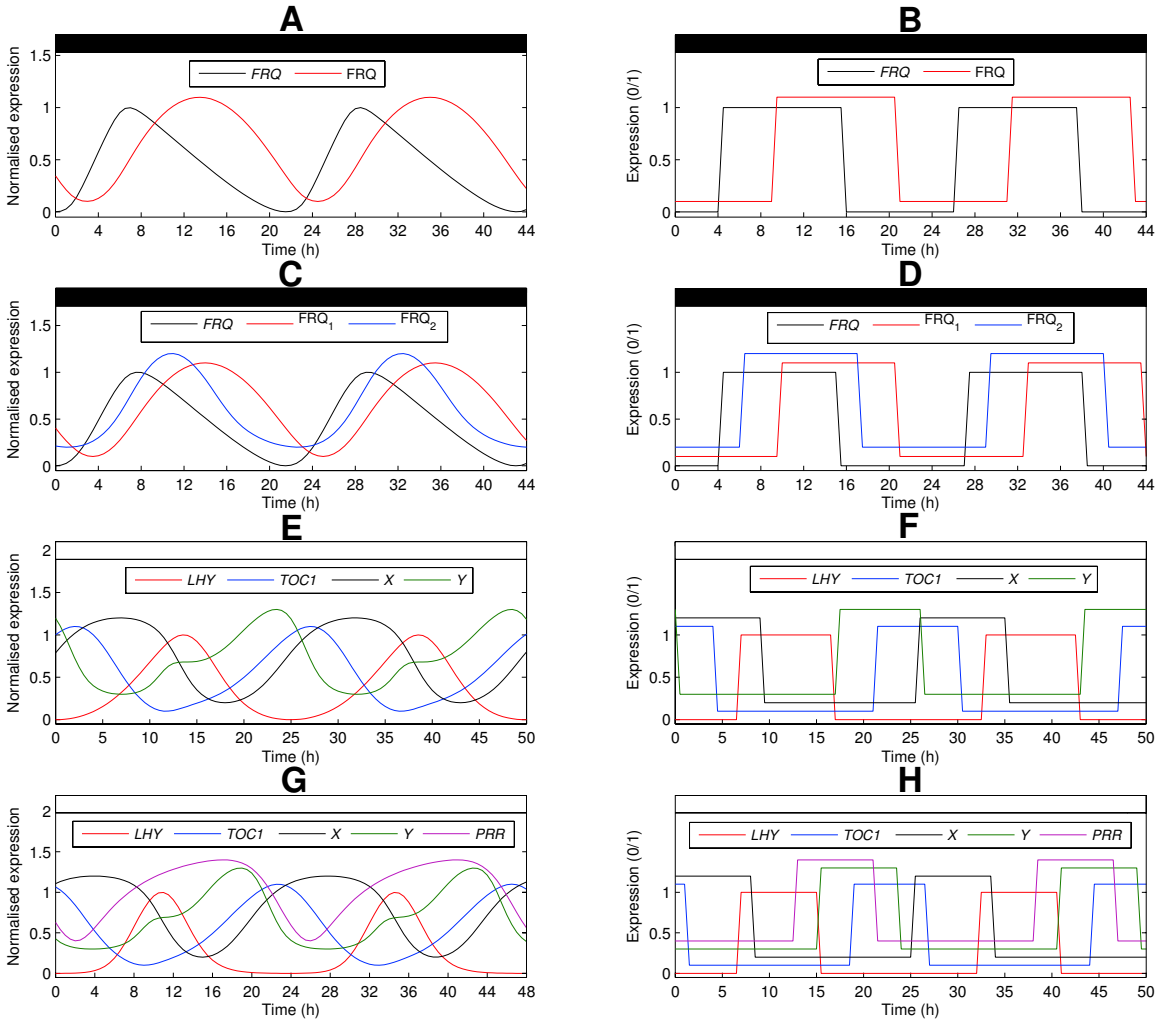

Figure S3: Time series for the differential equation (DE) and Boolean versions of the clock models in constant conditions. Two periods of each are plotted for comparison. **A**, **B**: 1-loop *Neurospora*; **C**, **D**: 2-loop *Neurospora*; **E**, **F**: 2-loop *Arabidopsis*; **G**, **H**: 3-loop *Arabidopsis*. DE time series (left panels) have been normalised to lie between 0 and 1 in order to facilitate comparison with the Boolean simulations (right panels). Different components within a model are slightly offset from one another so they can be distinguished more easily. The time step used for solving the Boolean models was 0.5h, equal to the data sampling interval.

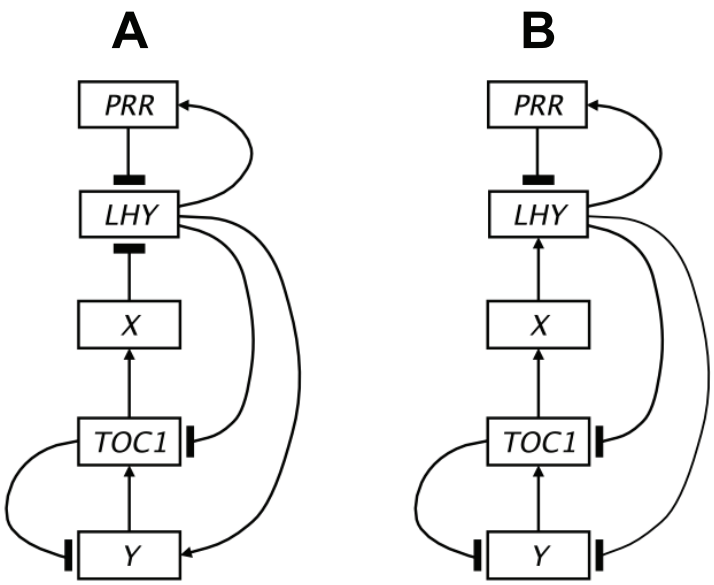

Figure S4: The logic configurations of the free-running 3-loop *Arabidopsis* model giving the best fits to experimental LUC data. **A**. The optimal configuration, G_{OPT} = (10101011011). **B**. The second highest ranking configuration, G_{DE} = (10011011011).

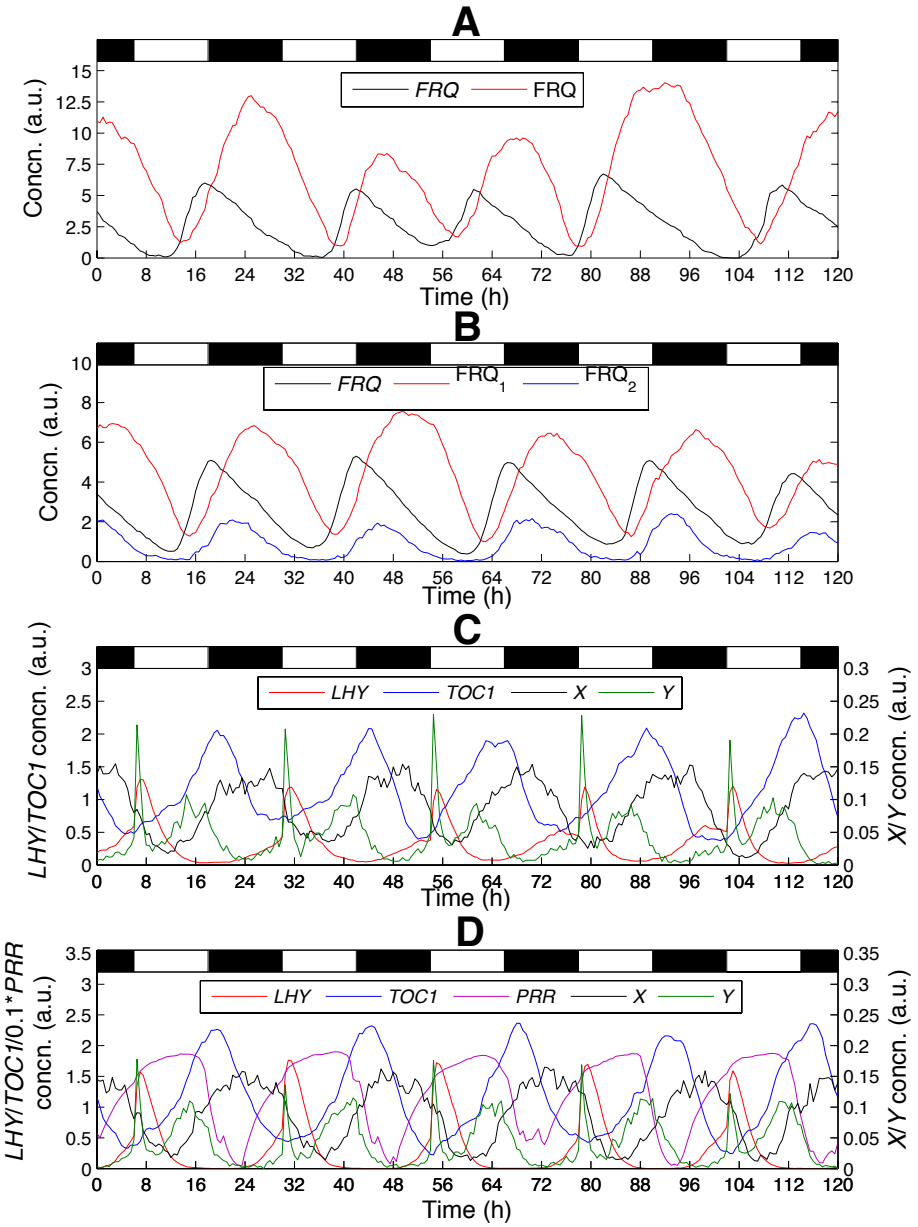

Figure S5: The synthetic LD data used to fit the Boolean clock models. **A**. 1-loop *Neurospora*. **B**. 2-loop *Neurospora*. **C**. 2-loop *Arabidopsis*. **D**. 3-loop *Arabidopsis*: *PRR* expression has been scaled to fit on the same axes as the other genes. Time series were generated for each model using the Gillespie algorithm, as described in the Methods section. White and black bars denote lights-on and lights-off respectively. The sampling interval is 0.5h.

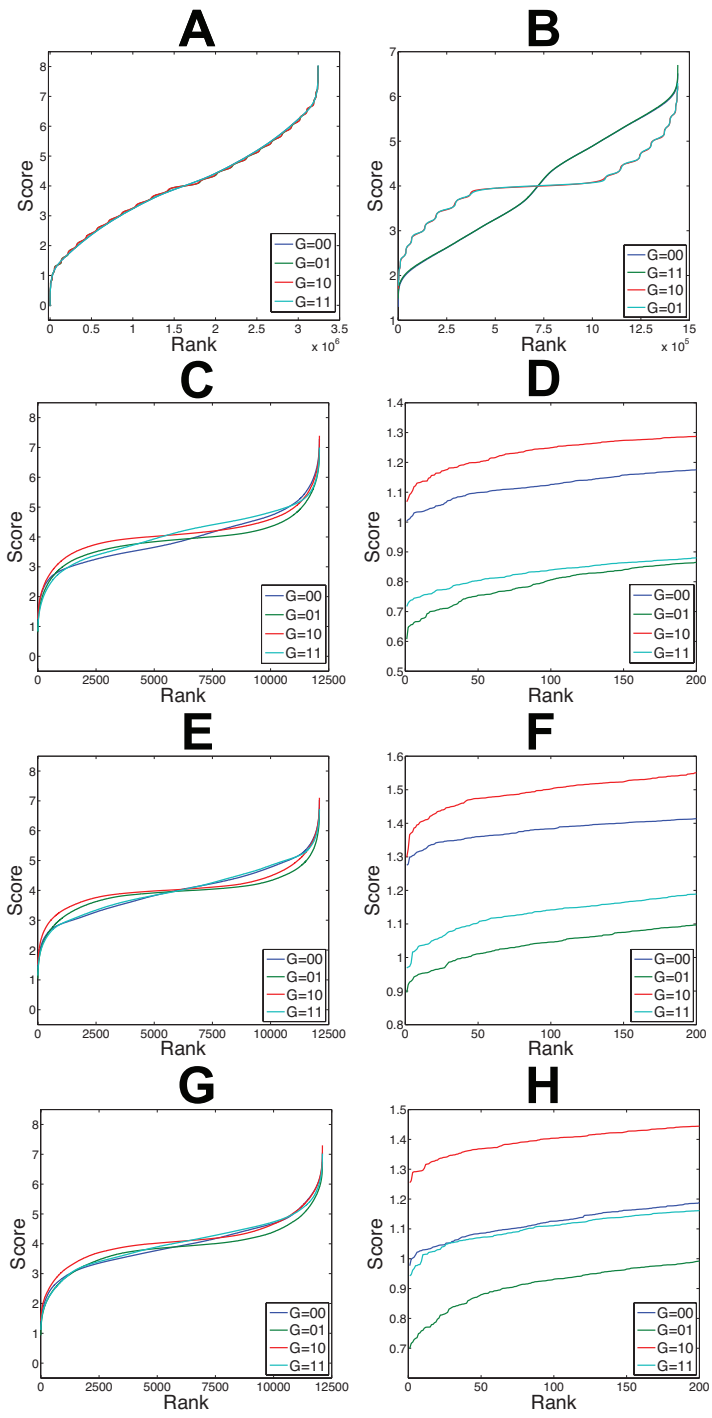

Figure S6: Degeneracies in the cost function for the 1-loop *Neurospora* model. **A**. 4-fold degeneracy in data with near equivalent durations of ON and OFF states. (Sinusoidal synthetic data with discretisation thresholds in the range [20%, 80%] of the minimum to maximum difference). **B**. 2-fold degeneracy in data with significantly imbalanced durations of ON and OFF states. (Same data with discretisation thresholds in the ranges [0%, 20%] and [80%, 100%] of the minimum to maximum difference). **C**-**H**. Pairs of figures showing global costs (left panels) and the lowest 200 costs (right panels) for each LC, obtained using increasingly noisy synthetic data. The degeneracy lifts for increasing noise, with the DE LC, $G = (01)$, separating from the others to emerge clearly as the configuration yielding the lowest cost.