Please cite this article in press as: Troein et al., Weather and Seasons Together Demand Complex Biological Clocks, Current Biology (2009), doi:10.1016/j.cub.2009.09.024

Current Biology *19*, 1–4, December 1, 2009 DOI 10.1016/j.cub.2009.09.024

Report

Weather and Seasons Together Demand Complex Biological Clocks

Carl Troein,[1,4] James C.W. Locke,[2,4] Matthew S. Turner,[3] and Andrew J. Millar[1,*]

[1]School of Biological Sciences, University of Edinburgh and Centre for Systems Biology at Edinburgh, Edinburgh EH9 3JR, UK

[2]Division of Biology and Department of Applied Physics, California Institute of Technology, M/C 114-96, Pasadena, CA 91125, USA

[3]Department of Physics, University of Warwick, Coventry CV4 7AL, UK

Summary

The 24-hour rhythms of the circadian clock [1] allow an organism to anticipate daily environmental cycles, giving it a competitive advantage [2, 3]. Although clock components show little protein sequence homology across phyla, multiple feedback loops and light inputs are universal features of clock networks [4, 5]. Why have circadian systems evolved such a complex structure? All biological clocks entrain a set of regulatory genes to the environmental cycle, in order to correctly time the expression of many downstream processes. Thus the question becomes: What aspects of the environment, and of the desired downstream regulation, are demanding the observed complexity? To answer this, we have evolved gene regulatory networks in silico, selecting for networks that correctly predict particular phases of the day under light/dark cycles. Gradually increasing the realism of the environmental cycles, we have tested the networks for the minimal characteristics of clocks observed in nature: oscillation under constant conditions, entrainment to light signals, and the presence of multiple feedback loops and light inputs. Realistic circadian gene networks are found to require a nontrivial combination of conditions, with seasonal differences in photoperiod as a necessary but not sufficient component.

Results and Discussion

Evolving Clock Networks

Starting from randomly connected networks of genes (Figure 1A), we have used a genetic algorithm to create clock networks in which one gene is designated to be expressed just after dawn and another just before dusk. This pattern exemplifies the well-characterized rhythmic profiles of core circadian clock genes, such as *Per1* and *Per2* in the mammalian suprachiasmatic nucleus or *PRR9* and *GI* in *Arabidopsis* [1]. A fitness function measures how well the network times expression of the dawn and dusk genes. Our approach differs from earlier work, which sought specifically for oscillations in constant conditions [6–8]. These works demonstrated that it is possible to evolve simple networks that oscillate and can be entrained to a light/dark cycle. We now use this technique to address the fundamental question of which properties of the environment are required to evolve the complex circadian networks found in nature.

To probe the role of the environmental input, we evolved networks under a range of light conditions. The most basic was alternating 12 hours of light and darkness (LD 12:12), and we extended this in two directions: multiple photoperiods and noise in the timing of the light signal. The former mimics seasonal differences, hypothesized to be important for the emergence of complex clocks [9], whereas the latter represents weather and other stochastic effects on the system. The effects of molecular noise on circadian clocks have been studied extensively [10–14], showing that simple one-loop oscillators can be robust to molecular noise, given the correct parameter choices. In this study, we focused on the effect of environmental noise on circadian clock evolution. To compare the idealized scenarios with natural conditions, we also evolved networks against a year-long time series of environmental radiometry data from Harvard Forest [15], where dawn and dusk change gradually and the light intensity fluctuates with the weather.

The networks were modeled as delay differential equations with parameters for light activation and for the signs, strengths, and timescales of gene-gene interactions. The choice of delays over mass action kinetics greatly reduces the number of parameters without being incompatible with biological systems [16–19]. For computational tractability, we limited the networks to no more than four genes. This limit was selected to allow a wide range of interlocking loop structures, comparable to the complexity of mechanistic circadian clock models. Over 10^8 network architectures were possible with four genes.

Network Analysis

The goal of using a genetic algorithm to optimize the topology and parameters was to create an ensemble of realistic networks. By strongly selecting for correct dawn and dusk gene expression, we removed most of the generated networks from further analysis. The absolute fitness and fitness distribution of the solutions varied significantly among scenarios, reflecting the challenges of the different environments and making it inappropriate to apply a single fitness threshold across scenarios. The 50 best performing solutions, out of 5,000 evolved, were therefore analyzed for each scenario. In general, biological networks might contain interactions that slightly increase fitness without being integral to function, so for the analysis of network structure, we exposed the functional network cores by iteratively removing the least important regulatory interaction or light input, stopping when the fitness would drop below 95% of its original value. The cores of the best performing networks are shown in Figures 1B–1F. For the single-photoperiod scenarios, the networks shown are representative of the 50 best solutions. What has evolved is a simple light-driven on/off switch for the dawn gene—an incoherent feed-forward loop with light as its input—with an additional delay for the dusk gene.

Figure 2 gives a summary of the evolved network structures and any sustained circadian oscillations. The simplest LD

*Correspondence: andrew.millar@ed.ac.uk

[4]These authors contributed equally to this work

Please cite this article in press as: Troein et al., Weather and Seasons Together Demand Complex Biological Clocks, Current Biology (2009), doi:10.1016/j.cub.2009.09.024

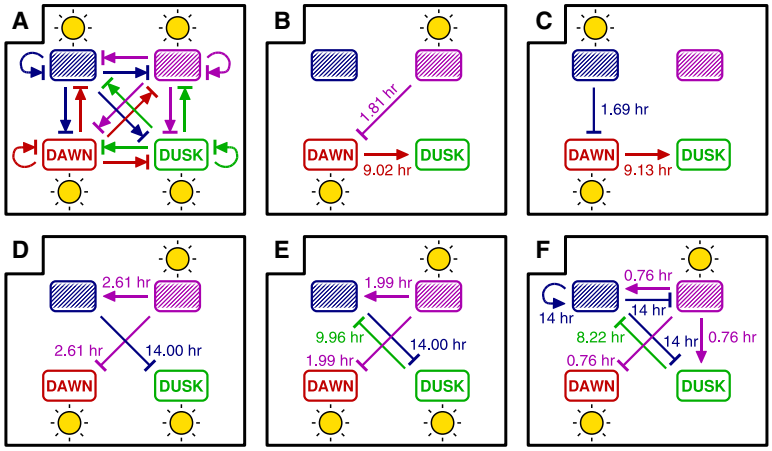

Figure 1. The Network Model

(A) The general form of the four-gene networks that we considered as candidates for generating circadian rhythms. Gene regulatory interactions may be positive, negative, or absent, and genes may be activated by light.

(B–F) The highest scoring network for five scenarios: (B) one photoperiod, (C) one photoperiod with noise, (D) multiple photoperiods, (E) multiple photoperiods with noise, and (F) one year of radiometry data. Gene interactions are shown with signs and delay times, and yellow suns denote light-activated expression. The designated dawn- and dusk-tracking genes are marked; the other two genes are interchangeable, so (B) and (C) are equivalent architectures. For both single-photoperiod scenarios (B and C), the networks shown represent the architecture of about 80% of the 50 best solutions (data not shown). The other 20% were functionally very similar, only replacing the positive regulation with a double negative. For multiple photoperiods (D), about 30% of the solutions looked like the one shown, whereas the last two scenarios did not use any one architecture for more than three of the solutions. The functional network cores were exposed by pruning of unimportant interactions (see text).

12:12 conditions only selected for delayed light responses, never oscillators, regardless of whether noise was added to the input. Extending the basic fitness function to multiple photoperiods had relatively little effect. The networks evolved few or no feedback loops, and circadian oscillations remained unlikely. In this scenario alone, we saw evidence of a tradeoff between light inputs and feedback loops, showing that under some circumstances, additional inputs are an alternative to increased structural complexity. However, combining multiple photoperiods with environmental noise eliminated that alternative strategy. Instead, the addition of noise led to a sharp increase in the number of feedback loops and in the probability of obtaining a circadian clock. Strikingly, networks faced with real environmental variations (Figure 3) evolved even more loops and light inputs and were most likely to exhibit circadian oscillations. Only in this scenario was the light level noisy during the day. Thus noise in the duration and level of the entraining light input signal appeared to favor greater complexity in the networks that timed gene expression.

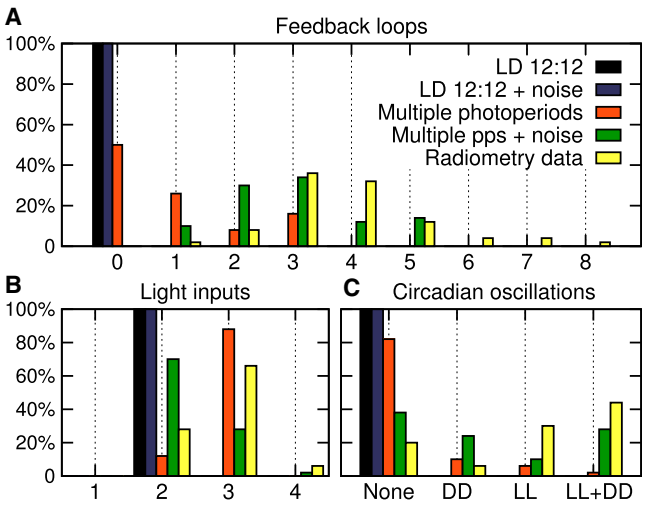

Figure 2. Complexity in Clock Networks Evolved under Different Environmental Input

The distribution of the number of feedback loops (A) and light inputs (B) in the functional cores of the top 50 networks from each scenario and the fraction of the networks that exhibit circadian oscillations only in constant light, darkness, or both (C). Increasingly realistic conditions led to more feedback loops, light inputs, and oscillations. The large numbers of light inputs selected under multiple photoperiods are discussed in the text.

Conclusions

A hallmark of circadian regulation is the ability to robustly adjust to different photoperiods despite unpredictable variations in temperature, light intensity, and other environmental parameters. By evolving systems in silico, we have explored the interactions between functional requirements on the timing of gene expression and robustness to noise in order to identify factors that can explain the ubiquity of multiloop circadian clocks. We have shown that seasonally changing photoperiods alone are insufficient to select for network complexity in a circadian system that can anticipate environmental transitions. However, when coupled with environmental noise, varying photoperiod strongly selects for complexity and gives rise to circadian clocks with multiple feedback loops and multiple light inputs, just as observed in nature.

Experimental Procedures

Network Model

The networks that we evolved are illustrated in Figure 1A. Transcription can be light activated, and genes might activate or repress the transcription of themselves and others. Posttranscriptional processes (including translation) give a discrete time delay of between 15 min and 14 hr. Following the time-averaged statistical treatment of Shea and Ackers [20], we modeled the system by four delay differential equations, each taking the form

$$\frac{dG_i}{dt} = S_i \frac{B_i + \Theta L_i o_{iL} + \sum_{j=1}^{4} a_{ij} o_{ij} \left(\frac{G_j(t-T_j)}{k_{ij}}\right)^2}{1 + B_i + \Theta L_i o_{iL} + \sum_{j=1}^{4} o_{ij} \left(\frac{G_j(t-T_j)}{k_{ij}}\right)^2} - D_i G_i(t),$$

where $G_i(t)$ is the level of gene i at time t, S_i its maximum transcription rate, B_i its basal expression level, and D_i its decay rate. Gene interactions are defined by the parameters o_{ij}, $a_{ij} \in \{0,1\}$. When $o_{ij} = 1$, there is repression ($a_{ij} = 0$) or activation ($a_{ij} = 1$) of gene i by gene j, with strength k_{ij} and time delay T_j. Similarly, if $o_{iL} = 1$, then light activates gene i with strength L_i when the entrainment signal $\Theta > 0$. The Hill coefficients for gene-gene interactions are fixed at 2. This model of a genetic network is highly simplified but nonetheless captures a wide range of network dynamics.

Fitness Function

Given parameter values and the input signal $\Theta(t)$, the G_i are determined as functions of time. The fitness score is based on the expression of one

Please cite this article in press as: Troein et al., Weather and Seasons Together Demand Complex Biological Clocks, Current Biology (2009), doi:10.1016/j.cub.2009.09.024

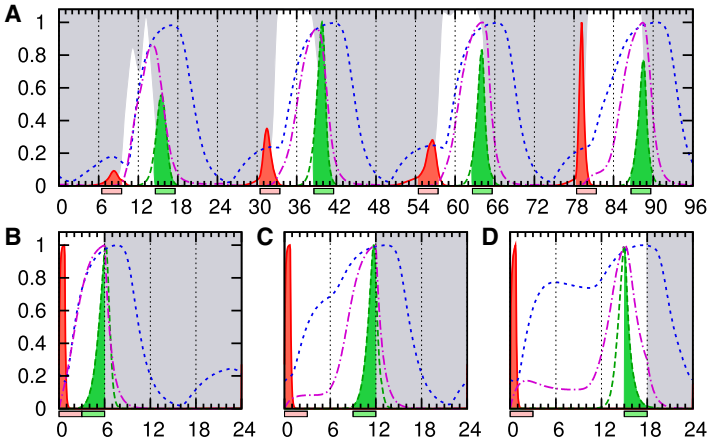

Figure 3. Network Dynamics with Real Environmental Input

Examples of network dynamics for the network of Figure 1F.

(A) A small part of the Harvard Forest radiometry data that the network was evolved against, and the corresponding gene expression time course, normalized to unit maximum. Periods of darkness are represented by gray shading. The target expression windows are indicated by red and green boxes for the dawn and dusk gene, respectively. The gene traces are plotted in the same colors as in Figure 1, with the parts matching the target expression windows shaded in red and green. Time $t = 0$ is midnight, not dawn, because there is no well-defined zeitgeber period and phase.

(B–D) The day length varies with the season between about 9 and 15 hr, but the network can be entrained to light/dark (LD) cycles with a wider range of photoperiods: LD 6:18 (B), LD 12:12 (C), and LD 18:6 (D). In (B)–(D), $t = 0$ is dawn.

gene in a 3 hr time window after dawn and of another gene in a similar window before dusk. The fitness for a single simulated day is then

$$f^{-1} \propto \left(\frac{\int_l^{l+3} G_1 dt}{\int_0^{24} G_1 dt}\right)^{-1} + \left(\frac{\int_{d-3}^{d} G_2 dt}{\int_0^{24} G_2 dt}\right)^{-1} + 0.01 \sum_{i=1}^{2} \left(\min\left(\frac{\int_0^{24} G_i dt}{1000}, 1\right)\right)^{-1} + 0.001 \left(\sum_{i,j} o_{ij} + \sum_i o_{iL}\right)^{-1},$$

normalized such that $0 \leq f \leq 1$. The first two terms describe the expression of the dawn and dusk genes in the time windows, relative to their totals, whereas the third term discourages very low expression levels. The last term is a small penalty on superfluous connections and light inputs, which mostly affects the simplest scenarios where the fitness differences between solutions are small. Without this term, feedback loops appeared in many networks even for the single-photoperiod scenarios, where they were not required for near perfect scores.

Simulations

To evaluate the fitness function for a given parameter set and light input signal, we implemented a delay differential equation solver in C++ using a fourth-order ordinary differential equation solver from the GNU Scientific Library (GSL) [21]. Hermite interpolation of the values and derivatives of the variables at the time points visited by the variable step-length ordinary differential equation (ODE) solver were used to provide system history for the delay terms and to evaluate the integrals of G_i. Each parameter set was thus always accompanied by its recent history, including current variable values. Simulations proceeded one day at a time, failing (reporting negative fitness) if more than 10^4 time steps were needed. Following any change, the system was converged toward a limit cycle for up to 20 days of identical light input, terminating early if end-of-day state or fitness score converged to within a 10^{-4} relative difference between several consecutive days. As a fallback, the worst fitness score of the last 15 days was reported.

Multiple Photoperiods and Noise

For multiphotoperiod scenarios, we used nine photoperiods between LD 6:18 and LD 18:6. The state was converged (as described above) following every photoperiod change. In scenarios with noise, the system ran for 24 days with dusk at nominal dusk ± 2 hr (flat distribution). The total fitness was the harmonic mean over the individual days. In the environmental data scenario, the system was converged against the first day of data, then simulated for a further 365 days. The input signal came from Harvard Forest data set HF102 (available at http://harvardforest.fas.harvard.edu/), specifically the hourly measurements of total incoming radiation for the year 2000. An arbitrary transformation was needed to give a level near 0 at night and saturated at 1 on sunny days. We used $\Theta = 0.5\tanh(x/30 - 2.5) + 0.5$ and interpolated between data points by a nonovershooting cubic spline (Figure 3). Nominal dawn and dusk at Harvard Forest, needed for the fitness scores, were computed using the date_sun_info function of the PHP programming language.

Genetic Algorithm

To evolve the networks, we used a real-coded genetic algorithm [22]. Our particular algorithm is described in detail in the Supplemental Data available online, and we give a brief summary here. In each generation, the bottom tenth of the individuals in a population of 50 parameter sets were replaced through cloning (including mutation of one or more parameters through multiplication by a random factor) or recombination (with new parameter values drawn from the vicinity of the two parents' values). The runs lasted between 1,500 and 25,000 generations, stopping when fitness could not be improved. Similar results were obtained from a different genetic algorithm in a separate implementation.

Circadian Oscillations

To test a network for circadian oscillations, we simulated the system for 10 days following entrainment in LD 12:12, switching to constant conditions in the first day. The expression levels of the dawn and dusk genes for days 4–8 were analyzed with fast Fourier transform nonlinear least squares (FFT-NLLS) [23] at confidence level 0.95. If any component with period between 15 and 35 hr was found, we considered the network to be a circadian oscillator. Figure S1 of the Supplemental Data shows the period and amplitude of the oscillations and that these are affected by the functional criteria used to evolve the networks. To remove very weakly oscillating networks, we required the root mean square (RMS) distance between the time course and a detrended version of the same time course to be at least 10% of the mean level for at least one gene.

Supplemental Data

Supplemental Data include Supplemental Experimental Procedures and one figure and can be found online at http://www.cell.com/current-biology/supplemental/S0960-9822(09)01704-7.

Acknowledgments

A.J.M. gratefully acknowledges the hospitality of L. Serrano and the Center for Genomic Regulation, Barcelona during revision of the manuscript. C.T. was supported by ANR/BBSRC grant BB/F005466 and the Human Frontiers Science Program. J.C.W.L. was supported by a Sainsbury studentship from the Gatsby Charitable Foundation. Computations were performed with the support of the Edinburgh Compute and Data Facility and the Centre for Scientific Computing at the University of Warwick.

Received: July 12, 2009
Revised: August 31, 2009
Accepted: September 2, 2009
Published online: October 8, 2009

References

1. Dunlap, J.C., Loros, J.J., and Decoursey, P.J. (2003). Chronobiology: Biological Timekeeping (Sunderland, MA: Sinauer).
2. Ouyang, Y., Andersson, C.R., Kondo, T., Golden, S.S., and Johnson, C.H. (1998). Resonating circadian clocks enhance fitness in cyanobacteria. Proc. Natl. Acad. Sci. USA *95*, 8660–8664.
3. Dodd, A.N., Salathia, N., Hall, A., Kevei, E., Toth, R., Nagy, F., Hibberd, J.M., Millar, A.J., and Webb, A.A. (2005). Plant circadian clocks increase

photosynthesis, growth, survival, and competitive advantage. Science *309*, 630–633.

4. Stelling, J., Sauer, U., Szallasi, Z., Doyle, F.J., 3rd, and Doyle, J. (2004). Robustness of cellular functions. Cell *118*, 675–685.
5. Rand, D.A., Shulgin, B.V., Salazar, D., and Millar, A.J. (2004). Design principles underlying circadian clocks. J. R. Soc. Interface *1*, 119–130.
6. François, P., and Hakim, V. (2004). Design of genetic networks with specified functions by evolution in silico. Proc. Natl. Acad. Sci. USA *101*, 580–585.
7. Wagner, A. (2005). Circuit topology and the evolution of robustness in two-gene circadian oscillators. Proc. Natl. Acad. Sci. USA *102*, 11775–11780.
8. Rodrigo, G., Carrera, J., and Jaramillo, A. (2008). Computational design and evolution of the oscillatory response under light–dark cycles. Biochimie *90*, 888–897.
9. Merrow, M., Spoelstra, K., and Roenneberg, T. (2005). The circadian cycle: Daily rhythms from behaviour to genes. EMBO Rep. *6*, 930–935.
10. Forger, D.B., and Peskin, C.S. (2005). Stochastic simulation of the mammalian circadian clock. Proc. Natl. Acad. Sci. USA *102*, 321–324.
11. Gonze, D., and Goldbeter, A. (2006). Circadian rhythms and molecular noise. Chaos *16*, 026110.
12. Gonze, D., Halloy, J., and Goldbeter, A. (2002). Robustness of circadian rhythms with respect to molecular noise. Proc. Natl. Acad. Sci. USA *99*, 673–678.
13. Vilar, J.M.G., Kueh, H.Y., Barkai, N., and Leibler, S. (2002). Mechanisms of noise-resistance in genetic oscillators. Proc. Natl. Acad. Sci. USA *99*, 5988–5992.
14. Chabot, J.R., Pedraza, J.M., Luitel, P., and van Oudenaarden, A. (2007). Stochastic gene expression out-of-steady-state in the cyanobacterial circadian clock. Nature *450*, 1249–1252.
15. Moore, K.E., Fitzjarrald, D.R., Sakai, R.K., Goulden, M.L., Munger, J.W., and Wofsy, S.C. (1996). Seasonal variation in radiative and turbulent exchange at a deciduous forest in central Massachusetts. J. Appl. Meteorol. *35*, 122–134.
16. Smolen, P., Baxter, D.A., and Byrne, J.H. (2002). A reduced model clarifies the role of feedback loops and time delays in the Drosophila circadian oscillator. Biophys. J. *83*, 2349–2359.
17. Smolen, P., Hardin, P.E., Lo, B.S., Baxter, D.A., and Byrne, J.H. (2004). Simulation of Drosophila circadian oscillations, mutations, and light responses by a model with VRI, PDP-1, and CLK. Biophys. J. *86*, 2786–2802.
18. Akman, O.E., Locke, J.C.W., Tang, S., Carré, I., Millar, A.J., and Rand, D.A. (2008). Isoform switching facilitates period control in the Neurospora crassa circadian clock. Mol. Syst. Biol. *4*, 164.
19. Lema, M.A., Golombek, D.A., and Echave, J. (2000). Delay model of the circadian pacemaker. J. Theor. Biol. *204*, 565–573.
20. Ackers, G.K., Johnson, A.D., and Shea, M.A. (1982). Quantitative model for gene regulation by lambda phage repressor. Proc. Natl. Acad. Sci. USA *79*, 1129–1133.
21. Galassi, M., Davies, J., Theiler, J., Gough, B., Jungman, G., Alken, P., Booth, M., and Rossi, F. (2009). GNU Scientific Library Reference Manual, Third Edition (Bristol: Network Theory).
22. Herrera, F., Lozano, M., and Verdegay, J.L. (1998). Tackling real-coded genetic algorithms: Operators and tools for behavioural analysis. Artif. Intell. Rev. *12*, 265–319.
23. Plautz, J.D., Straume, M., Stanewsky, R., Jamison, C.F., Brandes, C., Dowse, H.B., Hall, J.C., and Kay, S.A. (1997). Quantitative analysis of Drosophila period gene transcription in living animals. J. Biol. Rhythms *12*, 204–217.