

Identification and Modeling of Co-Rhythmic Genes from Micro-array Time Series Data

Wenxue Wang^{†,*}, Thanura Elvitigala[‡], Jana Stockel[§], Himadri B. Pakrasi[§], Bijoy K. Ghosh[†]

[†]Department of Mathematics and Statistics,
Texas Tech University, Lubbock, TX, USA

[‡]Department of Electrical and Systems Engg.,
Washington University, Saint Louis, MO, USA

[§]Biology Department,
Washington University, Saint Louis, MO, USA

*E-mail: wenxue.wang@ttu.edu

Extended Abstract

‘Circadian Rhythm’ is a biological phenomenon observed in a large number of organisms ranging from unicellular bacteria to human beings. The underlying biochemical mechanism is also understood for many of the organisms to a varying degree of details. It is unclear, however, how a circadian clock controls various different metabolic processes. In particular, it is of importance to understand if only one clock is enough or if multiple clocks are required. In this paper, transcriptome data from *Cyanothece*, a photosynthetic cyanobacteria, has been analyzed for the purpose of discovering genes whose expressions are rhythmically close (co-rhythmic). Subsequently we study if these rhythms can be modeled, up to phase, using a cascade of three phase oscillators. One of the phase oscillator in the network is derived from the model of a ‘limit cycle oscillator’ using KaiC protein (the master clock). We conclude that ‘Circadian Rhythms in *Cyanothece* transcriptome data can be dynamically modeled up to phase using a single master clock derived from limit cycle oscillator using KaiC protein cascaded with a pair of interconnected phase oscillators’. Biologically substrates of the phase oscillators are presently unknown.

Cyanothece is grown in ASP2-N medium under 12/12 hour light/dark cycle and samples are extracted every 4 hrs to generate the microarray data. The data are normalized using Lowess algorithm [3] and quality assessment is performed using T-test algorithm. Log ratio of the data from control and target is obtained at 11 time points and they show rhythmic patterns. This has been displayed in Fig. 1(A).

Discovering the co-rhythmic genes:

We conjecture that the rhythmic pattern in the microarray data can be modeled using a pair of fundamental frequencies (two sinusoids) with unknown amplitudes and phases. In Fig. 1(B) we show results of fitting the rhythmic component by sinusoids with different frequencies. In Fig. 1(C) we plot the error as a function of the frequency pair. An optimum global minima is noted. For each

¹This work is part of a Membrane Biology EMSL Scientific Grand Challenge project at the W. R. Wiley Environmental Molecular Sciences Laboratory, a national scientific user facility sponsored by the U.S. Department of Energy’s Office of Biological and Environmental Research (BER) program located at Pacific Northwest National Laboratory. PNNL is operated for the Department of Energy by Battelle. This work was also partially supported by funding from the NSF-FIBR program (EF0425749)

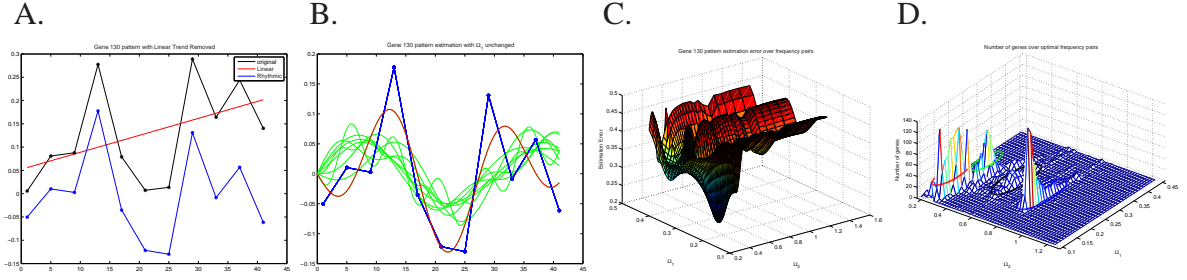


Figure 1: The log ratio pattern of micro array data and its estimation. A: The log ratio pattern of one gene (black) and its linear trend (red) and rhythmic component (blue). B: The rhythmic component (blue) and its estimates with two sinusoidal functions (green). The red curve is the best estimate with the optimum frequency pair. C: The estimation errors over frequency pairs. D: Clustering of genes based on the optimum frequency pairs.

gene with a rhythmic pattern of activity the corresponding optimum frequency pair is obtained and plotted in Fig. 1(D). A suitable clustering of this data set produces co-rhythmic genes.

Phase modeling of a master clock:

A suitable circadian oscillation in cyanothecae has been built using the phosphorylation/ dephosphorylation cycle of KaiC protein. Simulation of this cycle using a simplified dynamic model [2], shows convergence to a limit cycle as shown in Fig. 2(A). Phase activity of the KaiC oscillation, obtained after a suitable linear transformation, is modeled using a simple evolution equation described in [4] given by

$$\dot{\phi} = \omega + \sum_{m=1}^{\infty} (a_m \sin(m\phi) + b_m \cos(m\phi)), \quad (1)$$

where the parameters ω , a_m and b_m are obtained from the Fourier transform of the phase velocity (not shown here). Fig. 2(B) shows the phase activity of the KaiC model oscillation (blue) and the simulated phase activity using the evolution equation (1) (red). The phases are plotted in 2π modulus at the bottom.

Modeling with a network of three phase oscillators:

A typical rhythmic pattern of an expression data is modeled up to its phase. A network is constructed with three phase oscillators, shown in Fig. 2(C), with the associated dynamics given by

$$\begin{aligned} \dot{\phi}_0 &= \omega_0 + \sum_{m=1}^M (a_m \sin(m\phi_0) + b_m \cos(m\phi_0)) \\ \dot{\phi}_1 &= \omega_1 + \epsilon_1 \sin(k_0 k_1 \phi_1 - \phi_0 + \psi_1) + \epsilon_2 \sin(\phi_2 - k_1 \phi_1 + \psi_2) \\ \dot{\phi}_2 &= \omega_2 + \epsilon_3 \sin(k_0 \phi_2 - \phi_0 + \psi_3) + \epsilon_4 \sin(k_1 \phi_1 - \phi_2 + \psi_4), \end{aligned} \quad (2)$$

with $k_1 = \frac{\omega_2}{\omega_1}$ and $k_0 = \frac{\bar{\omega}}{\omega_2}$, where ϕ_0 is the phase of the master clock, ω_1, ω_2 are the optimum frequency pair and ϕ_1 and ϕ_2 are the phases of the two sinusoidal components obtained via pattern fitting of the micro array data. $\bar{\omega}$ is the average phase velocity of the master clock, which is the slope of the straight line that interpolates the phase activity of the master clock (the circadian phase pattern in this study). The above interconnected phase dynamics (2) is a variation of the well

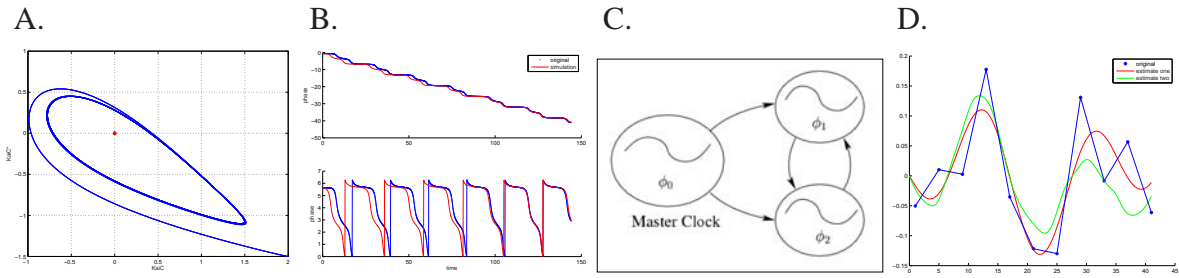


Figure 2: The phase activity of limit cycle oscillation of KaiC protein and oscillatory network simulation. A: The limit cycle oscillation of KaiC protein (KaiC* versus KaiC) with a linear transform. B: The phase activity of the KaiC model oscillation (blue) and the simulated phase activity using the evolution equation (1) (red). The phases are plotted in 2π modulus at the bottom. C: The oscillatory network including one master clock (ϕ_0) and two Kuramoto oscillators (ϕ_1 and ϕ_2). D: The genetic pattern with linear trend removed (blue) and its estimate (green) with phases, ϕ_1 and ϕ_2 , generated in the network (2). The red curve is the estimate with the optimum frequency pair.

known Kuramoto's models (see Kuramoto [1]). The rhythmic component of the micro array data can now be estimated using the sinusoidal functions of the phases ϕ_1 and ϕ_2 as follows:

$$\bar{g}(t) = \alpha_1 \sin(\phi_1(t)) + \alpha_2 \sin(\phi_2(t)),$$

and the result is shown in Fig. 2(D)

To conclude, we show that oscillations in micro array data can be reconstructed up to phase by a cascade of three phase oscillators. One of the phase oscillator is a model of the well known master clock obtained from the phosphorylation/dephosphorylation cycle of KaiC protein. Biological substrates of the other oscillators are presently unknown.

References

- [1] Y. Kuramoto, *Chemical Oscillations, Waves, and Turbulence*, Springer Verlag, 1984.
- [2] A. Mehra, C. I. Hong, M. Shi, J. J. Loros, J. C. Dunlap and P. Ruoff, "Circadian rhythmicity by autocatalysis", *PLoS Computational Biology*, 2(7): e96, pp. 816 – 823, 2006
- [3] J. Quackenbush, "Microarray data normalization and transformation", *Nature Genetics*, Vol. 32, pp. 496 – 501, 2002
- [4] P. A. Tass, *Phase Resetting in Medicine and Biology: Stochastic Modelling and Data Analysis*, Springer Verlag, 2006.