Proceedings of the 17th World Congress
The International Federation of Automatic Control
Seoul, Korea, July 6-11, 2008

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

Wenxue Wang * **Bijoy K. Ghosh** **

* *Texas Tech University, Department of Mathematics and Statistics, Lubbock, TX 79409 USA (e-mail: wenxue.wang@ttu.edu).*
** *Texas Tech University, Department of Mathematics and Statistics, Lubbock, TX 79409 USA (e-mail: bijoy.ghosh@ttu.edu)*

Abstract: 'Circadian Rhythm' is a biological phenomenon observed in a large number of organisms ranging from unicellular bacteria to human beings. 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.

Keywords: Modeling and Identification, Bioinformatics, Dynamics and Control, System Biology, Circadian Oscillation, Network

1. INTRODUCTION

'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 oscillators 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 the limit cycle oscillator using KaiC protein cascaded with a pair of interconnected phase oscillators'. Biologically substrates of the phase oscillators are presently unknown.

Cyanothece, a photosynthetic cyanobacteria, is grown in ASP2-N medium under 12/12 hour light/dark cycle. Under light conditions, cells are kept under 50 μE m^{-2}s^{-1} irradiance while it is 0 μE m^{-2}s^{-1} for the dark condition. Samples for mRNA are extracted every 4 hours, at 12 time points (hour 1, hour 5, through hour 45), to generate the micro array data. The data are associated with more than 6000 genes. A mixture of mRNA extractions from different time points has been used as reference channel. The data are normalized using Lowess algorithm and quality assessment is performed using T-test algorithm (see Quackenbush [2002]). Log ratios of the data from control signals and target signals are obtained at 12 time points and they show rhythmic patterns. The rhythmicity has been displayed in Fig. 1. In the figure, the black pattern is one example from more than 6000 log ratios of the normalized micro array series data. The rhythmic component (the blue curve) can be obtained by removing the linear trend in the original pattern. The rhythmic component can be reconstructed with two sinusoidal functions of a pair of optimum frequencies. A gene clustering strategy is proposed based on the optimum frequency pairs, instead of patten similarity.

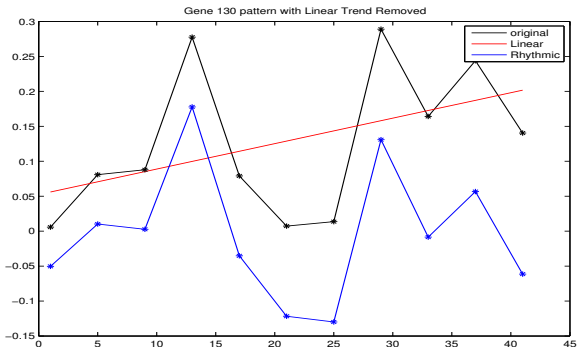

Fig. 1. The log ration pattern of micro data (black) and its linear trend (red) and rhythmic component (blue).

Circadian rhythms are a central feature of biological systems and are environmentally adaptive, and homeostatically regulated oscillators have been found in many organisms (cyanobacteria, fungi, and flies). Oscillation in KaiC phosphorylation

⋆ This 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).

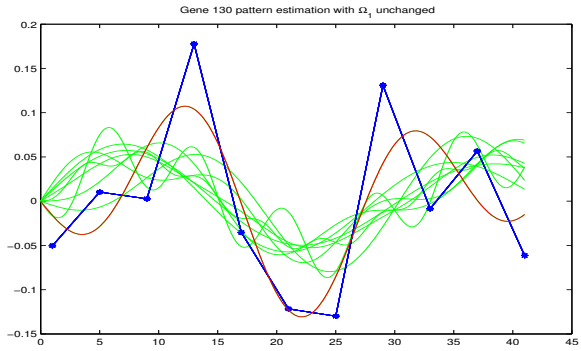

Fig. 2. The rhythmic component in the micro array expression data and its estimates with two sinusoidal functions over different pairs of frequencies. The red curve is the estimate with the optimum frequency pair.

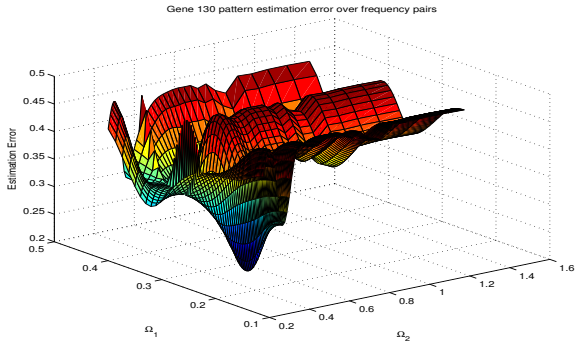

Fig. 3. The estimation errors over the pairs of frequencies. The optimum frequency pair is chosen that caused the smallest error and the optimum frequency pair makes best estimate (red curve in Fig. 2).

is a key regulator for the clock *in vivo*. A suitable circadian oscillation in cyanothece has been built using the phosphorylation/dephosphorylation cycle of KaiC protein. Simulation of this cycle using a dynamic model for the *in vitro* KaiC circadian clock (see Mehra et al. [2006]) shows convergence to a limit cycle. Phase activity of the KaiC oscillation, obtained after a suitable linear transformation, is modeled using a simple evolution equation described in Tass [2006].

The typical rhythmic pattern of micro array expression data can be modeled up to its phase. A oscillatory network is proposed with three phase oscillators. The oscillations in micro array series data are reconstructed up to phases by a cascade of three oscillators. One of the phase oscillator is a model of master clock obtained from phosphorylation/dephophorhylation cycle of KaiC protein. The other two oscillators are the model of the phase activities of two rhythmic components with the optimum frequency pairs in the micro array expression patterns and Kuramoto model has been utilized (see Kuramoto [1984]).

2. ESTIMATION OF MICRO ARRAY EXPRESSION PATTERNS WITH OPTIMUM FREQUENCY PAIR.

The log ratio patterns of 12 time points collected in micro array experiments show rhythmicity after the process of Lowess normalization and T-test. We conjecture that the rhythmic pattern in the micro array data can be modeled using a pair of fundamental frequencies (two sinusoids) with unknown amplitudes and phases. In this study we estimate the micro array pattern using linear combination of a linear trend and two sinusoidal functions with a pair of optimum frequencies described as follows:

$$g(t) = a + bt + \alpha_1 \sin(\omega_1 t + \theta_1) + \alpha_2 \sin(\omega_2 t + \theta_2)$$

where $a+bt$ is the linear component in the estimation, ω_1 and ω_2 are the frequencies of the rhythmic component, α_1 and α_2 are the amplitudes, θ_1 and θ_2 are the phases related to two sinusoidal components. Fig. 1 shows one example log ratio patten (black curve) that displays rhythmicity and its linear trend (red) and rhythmic component (blue). The part of linear trend can be obtained using linear regression. For the pairs $\{(t_i, g_i), i = 1, \cdots, N\}$, we can fit the curve $\{(t_i, g_i)\}$ using a linear function:

$$g(t) = a + bt$$

where the parameter a and b are chosen to minimize the cost function:

$$C(a,b) = \sum_{i=1}^{N} (g_i - a - bt_i)^2.$$

The least square method gives:

$$a = \frac{\sum g_i - b\sum t_i}{N}$$

$$b = \frac{N\sum g_i t_i - \sum g_i \sum t_i}{N\sum t_i^2 - (\sum t_i)^2}.$$

The rhythmic component in the micro array expression pattern then can be obtained by removing the linear trend:

$$\bar{g}(t_i) = g(t_i) - a - bt_i.$$

with obtained a and b (see Fig. 1). The rhythmic component can be estimated using a linear combination of two sinusoidal functions described as follows:

$$\bar{g}(t_i) = \alpha_1 \sin(\omega_1 t_i + \theta_1) + \alpha_2 \sin(\omega_2 t_i + \theta_2).$$

Given a pair of frequencies, (ω_1, ω_2), the parameters $\alpha_1, \alpha_2, \theta_1, \theta_2$ are chosen in such a way that the cost function is minimized:

$$E(\omega_1, \omega_2) = \min_{\alpha_1, \alpha_2, \theta_1, \theta_2} G(\alpha_1, \alpha_2, \theta_1, \theta_2)$$

where $G(\alpha_1, \alpha_2, \theta_1, \theta_2)$ is the cost function:

$$G(\alpha_1, \alpha_2, \theta_1, \theta_2) = \sum_{i=1}^{N} (\bar{g}(t_i) - \alpha_1 \sin(\omega_1 t_i + \theta_1) - \alpha_2 \sin(\omega_2 t_i + \theta_2))^2.$$

For the given frequency pair, (ω_1, ω_2), the optimal parameters $\alpha_1, \alpha_2, \theta_1, \theta_2$ are obtained using the gradient method:

$$\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \theta_1 \\ \theta_2 \end{bmatrix}_{k+1} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \theta_1 \\ \theta_2 \end{bmatrix}_{k} - \mu \nabla G(\alpha_1, \alpha_2, \theta_1, \theta_2)$$

where μ is the size of change in the vector $(\alpha_1, \alpha_2, \theta_1, \theta_2)$ and $\nabla G(\alpha_1, \alpha_2, \theta_1, \theta_2)$ is the gradient of the function $G(\alpha_1, \alpha_2, \theta_1, \theta_2)$:

$$\nabla G = \begin{bmatrix} \frac{\partial G}{\partial \alpha_1} \\ \frac{\partial G}{\partial \alpha_2} \\ \frac{\partial G}{\partial \theta_1} \\ \frac{\partial G}{\partial \theta_2} \end{bmatrix} = \begin{bmatrix} -2\sum_{i=1}^{N} \sin(\omega_1 t_i + \theta_1) e_i(\alpha_1, \alpha_2, \theta_1 \theta_2) \\ -2\sum_{i=1}^{N} \sin(\omega_2 t_i + \theta_2) e_i(\alpha_1, \alpha_2, \theta_1 \theta_2) \\ -2\alpha_1 \sum_{i=1}^{N} \cos(\omega_1 t_i + \theta_1) e_i(\alpha_1, \alpha_2, \theta_1 \theta_2) \\ -2\alpha_2 \sum_{i=1}^{N} \cos(\omega_2 t_i + \theta_2) e_i(\alpha_1, \alpha_2, \theta_1 \theta_2) \end{bmatrix}$$

where

$$e_i(\alpha_1, \alpha_2, \theta_1 \theta_2) = \bar{g}(t_i) - \alpha_1 \sin(\omega_1 t_i + \theta_1) - \alpha_2 \sin(\omega_2 t_i + \theta_2).$$

The frequency pairs, (ω_1, ω_2), are varied and the estimation errors over (ω_1, ω_2) have been computed . The optimum frequency pair is chosen that make the smallest errors. Fig. 2 shows the rhythmic component of one micro array expression pattern (blue) and its estimates with two sinusoidal functions over different pairs of frequencies (green smooth curves). The red smooth curve is

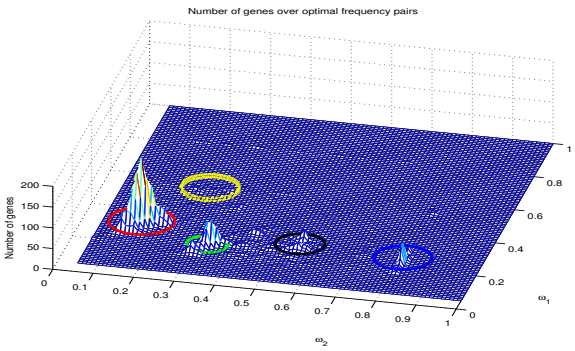

Fig. 4. The number of genes over optimum frequency pairs.

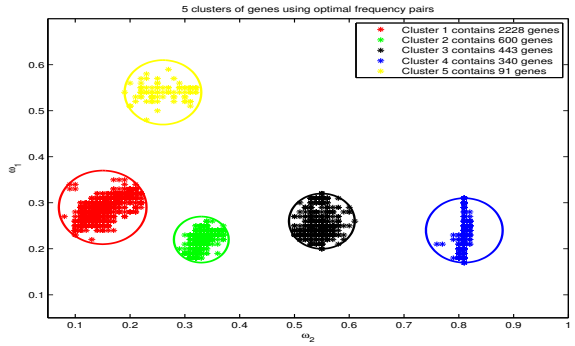

Fig. 5. Five clusters of genes are chosen based on optimum frequency pair.

the best estimate with the optimum frequency pair. The estimation error over frequency pairs is shown in Fig. 3. The frequency pair that cause the smallest estimation error is the optimum frequency pair that makes the best estimate (red curve in Fig. 2).

The genes that have the same optimum frequency pair are counted together. Fig. 4 shows the number of genes over the optimum frequency pair. The optimum frequency pairs provides a method of gene clustering based on the similarity of rhythmicity contained in the genetic expression patterns of more than 6000 genes. Fig. 5 shows four clusters of genes based on the optimum frequency pairs.

3. SIMULATION OF CIRCADIAN PHASE PATTERN OF KAIC PROTEIN PHOSPHORYLATION

Circadian rhythms are a central feature of biological systems and are environmentally adaptive. Homeopathically regulated oscillators have been found in many organisms (cyanobacteria, fungi, and flies). Oscillation in KaiC phosphorylation is a key regulator for the clock *in vivo*. KaiC is an enzyme with autokinase and autophosphatase activities. A synthetic, predictive and dynamic model for the *in vitro* KaiC phosphorylation activity has been built based on the self-amplifying response ('autocatalysis') of autophosphorylating kinases (see Mehra et al. [2006]), described as follows:

$$\begin{aligned}
\dot{x}_1 &= k_5 x_8 - k_1 x_1 x_3 - k_3 x_6 x_1 x_3 \\
\dot{x}_2 &= k_6 x_7 - k_4 x_6 x_2 \\
\dot{x}_3 &= k_7 x_4 - k_1 x_1 x_3 - k_3 x_6 x_1 x_3 \\
\dot{x}_4 &= k_6 x_7 - k_7 x_4 \\
\dot{x}_5 &= k_1 x_1 x_3 - k_2 x_5 \\
\dot{x}_6 &= k_2 x_5 + k_3 x_6 x_1 x_3 - k_4 x_6 x_2 \\
\dot{x}_7 &= k_5 x_8 - k_6 x_7 \\
\dot{x}_8 &= k_4 x_6 x_2 - k_5 x_8
\end{aligned}$$

where $x_1, \ldots, x_8$ represent KaiA, KaiB, KaiC, KaiC*, KaiAC, KaiAC*, KaiBC* and KaiABC* respectively. The model was described in details in Mehra et al. [2006]. Among these variables, x_3 and x_4, i.e., unphosphorylated KaiC protein and phosphorylated KaiC protein (KaiC*) are of interest in this study. Activities of KaiC and KaiC* are oscillatory and they converge to a limit cycle on the phase plane with proper initial conditions. Fig. 6 shows the oscillatory activities

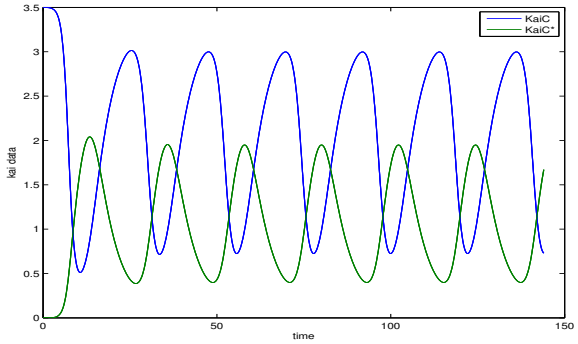

Fig. 6. The oscillatory activities of unphosphorylated KaiC and phosphorylated KaiC (KaiC*) in time.

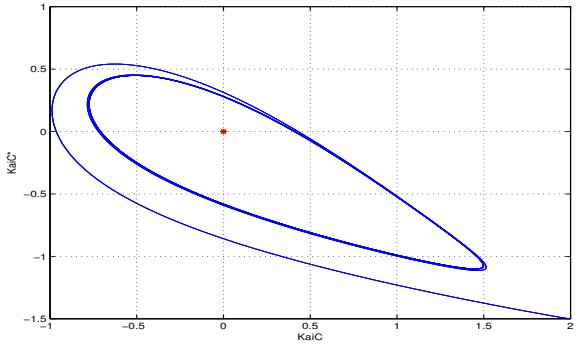

Fig. 7. An oscillatory behavior around the origin on the phase plane is obtained by taking a proper linear transformation.

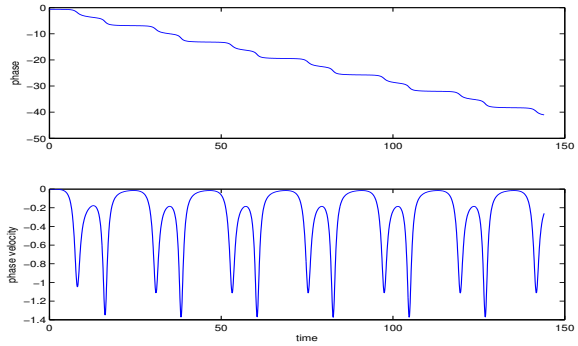

Fig. 8. The artificial phase pattern (top) and its velocity pattern (bottom) in time.

of KaiC and KaiC* in time. The activities of unphosphorylated KaiC and phosphorylated KaiC (KaiC*) forms a limit cycle oscillations. Taking a proper linear transformation of the oscillatory activities, we can have the oscillatory behavior around the origin on the phase plane. Then we can obtain an monotonic artificial phase activity of KaiC phosphorylation/dephosphrylation cycle. Fig. 7 shows the oscillatory behavior of KaiC and KaiC* protein around the origin on the phase plane obtained by taking a proper linear transformation, and Fig. 8 shows the artificial phase pattern of the KaiC phosphorylation/dephosphrylation cycle with the linear transformation and its velocity pattern in time. In Fig. 9 (top), the phase velocity pattern is plotted versus the phase pattern.

Since the oscillatory activities of KaiC and KaiC* proteins converge to a limit cycle, the phase pattern contains the clock information of the circadian regulator. One important and interesting problem is if the phase activity of KaiC phosphorylation/dephosphrylation oscillation can be modeled with some dynamical model. In this study we utilize the simple evolution equation described in Tass [2006]:

$$\dot{z} = (\alpha + i\omega)z - \beta z^2 z^* + \tilde{S}(z, z^*)$$

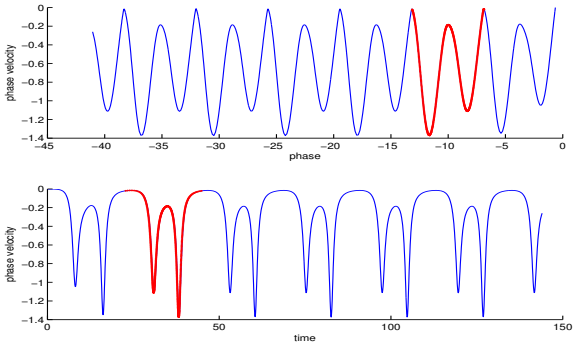

Fig. 9. The phase velocity is plotted versus the phase pattern (top), The section of phase velocity pattern (red part) has been used to compute the Fourier coefficients in the evolution equation (1).

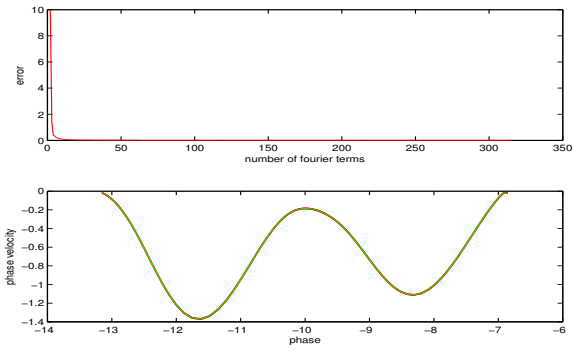

Fig. 10. The estimation of the part of phase velocity pattern using the Fourier Transform in phase with the coefficients obtained. On the top is the estimation error versus the number of Fourier transform coefficients and on the bottom is the estimate (green color) of the actual phase velocity (red color) using 30 pairs of Fourier transform coefficients.

where z is a complex variable, and z^* denotes the complex conjugate of z. ω is the eigenfrequency, α is a real parameter and β a nonnegative real parameter. $\tilde{S}(z,z^*)$, depending on z and z^*, represents a class of stimulation mechanisms which merely influence the phase. One special form of choice of $\tilde{S}$ is given by

$$\tilde{S}(z,z^*) = z\sum_{m=1}^{\infty}(\eta_m z^{m+1}z^* - \eta_m^*(z^*)^{m+1}z)$$

with complex coefficients

$$\eta_m = w_m + iv_m \qquad (w_m, v_m \in R).$$

Introduction of polar coordinates r and ϕ by inserting the hypothesis $z(t) = r(t)\exp(i\phi(t))$ gives the evolution equations of the amplitude r and the phase ϕ:

$$\begin{aligned}\dot{r} &= \alpha r - \beta r^3,\\ \dot{\phi} &= \omega + \textstyle\sum_{m=1}^{\infty}(2r^{m+2}w_m\sin(m\phi) + 2r^{m+2}v_m\cos(m\phi)).\end{aligned}$$

The stimulus $\tilde{S}(z,z^*)$ only acts on the oscillator's phase whereas the amplitude remains unaffected. For this reason the amplitude r relaxes towards the stable value $\sqrt{\alpha/\beta}$ irrespective of the initial value $r(0)$ and irrespective of the stimulus' action. So the evolution equation of the phase ϕ can be simplified into:

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

where a_m and b_m are real constant coefficients irrespective of the amplitude r.

In Fig. 9 (top), the plot of the phase velocity pattern versus the phase pattern shows that the phase velocity is a periodic function of the phase pattern. In the evolution equation of the phase(1), the derivative of phase is represented in the form of Fourier expansion in phase. These observations tell that the phase pattern can be reproduced with the evolution equation (1). To reproduce the

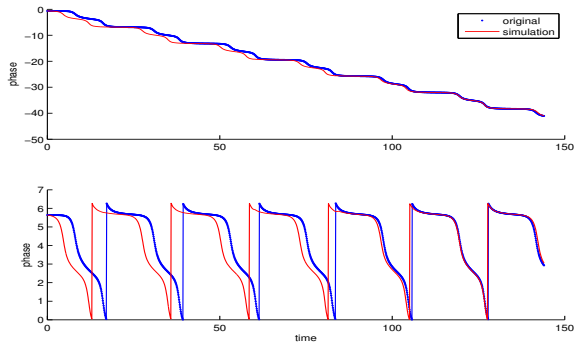

Fig. 11. The reproduction of the phase using the evolution equation (1) with the obtained Fourier coefficients. The blue curve is the original phase pattern and the red curve is the simulated phase pattern. The phases are plotted in 2π modulus at the bottom.

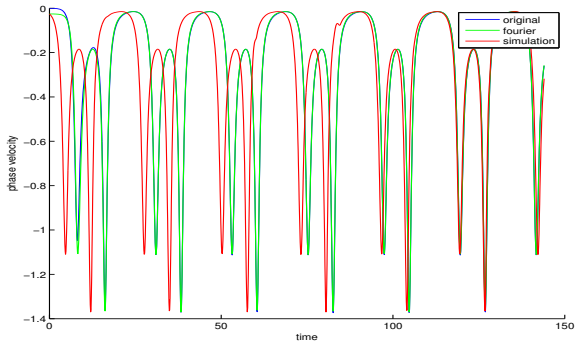

Fig. 12. Phase velocity pattern (in time) comparison: original phase velocity (blue), the estimate of phase velocity using the Fourier expansion (green), and the estimate of phase velocity using the evolution equation (red).

phase pattern using the evolution equation (1), the parameters in the model, ω, a_m and b_m have to be found. A period of phase velocity pattern has been used to compute those Fourier coefficients with DFT algorithm. Fig. 9 shows the section of phase velocity pattern (red) that is used to compute the coefficients. Fig. 10 shows the estimation of phase velocity pattern using the Fourier expansion in phase with the coefficients obtained. On the top is the estimation error versus the number of Fourier expansion coefficients and on the bottom is the estimate (green color) of the actual phase velocity (red color) using 30 pairs of Fourier coefficients. With the obtained Fourier coefficients, we can reproduce the phase pattern by simulating the evolution equation (1). Fig. 11 shows the simulation of phase pattern using the evolution equation, in which the blue curve is the original phase pattern and the red curve is the simulated phase pattern. Fig. 12 shows the comparison of original phase velocity and its estimates.

4. OSCILLATORY NETWORK WITH CIRCADIAN CLOCK OSCILLATOR

In Section 2, we studied the micro array expression pattern fitting using the linear combination of a linear trend and two sinusoidal components with an optimum frequency pair. In Section 3, we simulated the circadian activities of KaiC and KaiC* proteins using a dynamical model and the artificial phase activity of KaiC phosphorylation/dephosphorylation oscillation was reproduced using an phase model of evolution equation. A typical rhythmic pattern of an expression data can be modelled up to its phase. In this section, we propose a network of 3 oscillators that reproduce the circadian phase pattern and the phase patterns of two sinusoidal components in the micro array expression pattern fitting, shown in Fig. 13. Then the rhythmic component of micro array expression pattern with linear trend removed can be estimated with the sinusoids of the phase patterns of two sinusoidal components. One oscillator of the network is described by the evolution equation that reproduce master circadian clock and control the

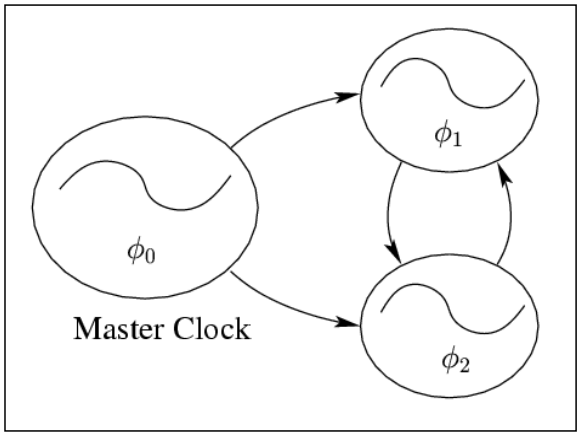

Fig. 13. The oscillatory network including one master clock (ϕ_0) and two Kuramoto oscillators (ϕ_1 and ϕ_2).

phases of sinusoidal components. The other two oscillators of the network are described by a modified Kuramoto model of two unit with extra terms which are the functions of master circadian phase. The oscillatory network is described as follows:

$$\begin{aligned}\dot{\phi}_0 &= \omega_0 + \textstyle\sum_{m=1}^{M}(a_m \sin(m\phi_0) + b_m \cos(m\phi_0)) \\ \dot{\phi}_1 &= \omega_1 + \varepsilon_1 \sin(k_0 k_1 \phi_1 - \phi_0 + \psi_1) \\ &\quad + \varepsilon_2 \sin(\phi_2 - k_1\phi_1 + \psi_2) \\ \dot{\phi}_2 &= \omega_2 + \varepsilon_3 \sin(k_0\phi_2 - \phi_0 + \psi_3) \\ &\quad + \varepsilon_4 \sin(k_1\phi_1 - \phi_2 + \psi_4),\end{aligned} \tag{2}$$

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 with $\omega_0 < 0$, ϕ_1 and ϕ_2 are the phases of the two sinusoidal components with natural frequencies, $\omega_1 > 0$, $\omega_2 > 0$, the optimum frequency pair obtained via pattern fitting of the micro array data. $\bar{\omega} < 0$ 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 Fig. 8). There exists a constraint, $\psi_4 = -\psi_2 = \frac{\psi_1 - \psi_2}{k_0}$. The connection parameters ε_i-s, $i = 1, \cdots, 4$, are positive. The above interconnected phase dynamics (2) is a variation of the well known Kuramoto's models (see Kuramoto [1984]). The rhythmic component of the micro array data now can be estimated using the sinusoidal functions of the phases ϕ_1 and ϕ_2 as follows:

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

Here we picked up one micro array expression pattern, as shown in Fig. 1, for example. Using the optimum frequency pair in the micro array pattern fitting obtained in Section 2 and the Fourier coefficients obtained in Section 3, and choosing values for other parameters in the dynamical model of oscillatory network, the phase patterns are simulated. Fig. 14 shows the master circadian phase pattern and two phase patterns for the sinusoidal components simulated with the model of oscillatory network (2). Fig. 15 shows the corresponding sinusoids of the phase patterns. Fig. 16 shows the estimation of the genetic pattern using sinusoids of the phases, $\phi_1(t)$ and $\phi_2(t)$, generated by the oscillatory network (green curve).

One important problem in the dynamical model (2) is the stability of the 'scaled' phase differences, $k_0k_1\phi_1 - \phi_0$ and $k_0\phi_2 - \phi_1$, that is, $k_0k_1\phi_1(t) - \phi_0(t)$ and $k_0\phi_2(t) - \phi_1(t)$ converge to some equilibrium points. Another phase difference in the model is $\phi_2 - k_1\phi_1$, which is dependent on the phase differences $k_0k_1\phi_1 - \phi_0$ and $k_0\phi_2 - \phi_1$. We start this analysis with a simpler model with those Fourier terms removed in the first equation, described as follows:

$$\begin{aligned}\dot{\phi}_0 &= \omega_0 \\ \dot{\phi}_1 &= \omega_1 + \varepsilon_1 \sin(k_0 k_1 \phi_1 - \phi_0 + \psi_1) \\ &\quad + \varepsilon_2 \sin(\phi_2 - k_1\phi_1 + \psi_2) \\ \dot{\phi}_2 &= \omega_2 + \varepsilon_3 \sin(k_0\phi_2 - \phi_0 + \psi_3) \\ &\quad + \varepsilon_4 \sin(k_1\phi_1 - \phi_2 + \psi_4),\end{aligned} \tag{3}$$

in which $\bar{\omega} = \omega_0$. Multiplying k_0k_1 on the second equation and k_0 on the third equation, we have:

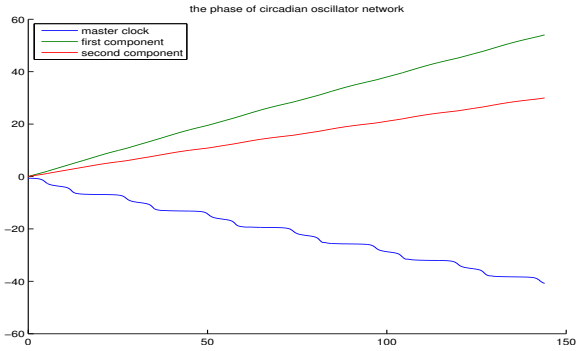

Fig. 14. The circadian phase pattern $\phi_0(t)$(blue) and two phase patterns, $\phi_1(t)$ and $\phi_2(t)$, for the sinusoidal components (green and blue) generated by simulating the oscillatory network (2)

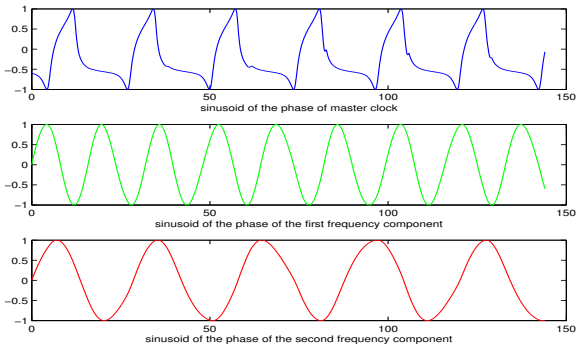

Fig. 15. The sinusoids of phase patterns shown in Fig. 14. The blue curve on the top is the sinusoid of the circadian phase $\sin(\phi_0(t))$. The other two curves (green and red) are the sinusoids of the phase patterns, $\sin(\phi_1(t))$ and $\sin(\phi_2(t))$, for the two sinusoidal components in the genetic pattern.

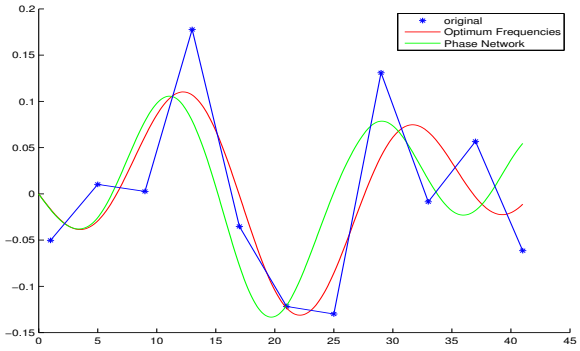

Fig. 16. Estimation of micro array expression pattern. The blue curve is the rhythmic component of the original micro array expression pattern, as shown in Fig. 1; the red curve is the estimate using two sinusoidal component with the optimum frequency pair; the green curve is the estimate using the sinusoids of the phases generated with the model of oscillatory network (2).

$$\begin{aligned}\dot{\phi}_0 &= \omega_0 \\ k_0k_1\dot{\phi}_1 &= k_0k_1\omega_1 + k_0k_1\varepsilon_1 \sin(k_0k_1\phi_1 - \phi_0 + \psi_1) \\ &\quad + k_0k_1\varepsilon_2 \sin(\phi_2 - k_1\phi_1 + \psi_2) \\ k_0\dot{\phi}_2 &= k_0\omega_2 + k_0\varepsilon_3 \sin(k_0\phi_2 - \phi_0 + \psi_3) \\ &\quad + k_0\varepsilon_4 \sin(k_1\phi_1 - \phi_2 + \psi_4)\end{aligned} \tag{4}$$

where $k_0k_1\omega_1 = k_0\omega_2 = \omega_0$ by the definitions of k_0 and k_1. Define $\eta_1 = k_0k_1\varepsilon_1, \eta_2 = k_0k_1\varepsilon_2, \eta_3 = k_0\varepsilon_3$, and, $\eta_4 = k_0\varepsilon_4$, then η_i-s, $i = 1, \cdots, 4$, are

negative because k_0 is negative. Define $\Phi_1 = k_0k_1\phi_1 - \phi_0 + \psi_1$ and $\Phi_2 = k_0\phi_2 - \phi_0 + \psi_3$, and by the constraint $\psi_4 = -\psi_2 = \frac{\psi_1-\psi_2}{k_0}$, we have $\phi_2 - k_1\phi_1 + \psi_2 = \frac{\Phi_2-\Phi_1}{k_0}$ and $k_1\phi_1 - \phi_2 + \psi_4 = \frac{\Phi_1-\Phi_2}{k_0}$. From (4) and the new variables defined above, we have differential equations on Φ_1 and Φ_2 as follows:

$$\begin{aligned} \dot{\Phi}_1 &= \eta_1 \sin(\Phi_1) + \eta_2 \sin(\tfrac{\Phi_2-\Phi_1}{k_0}) \\ \dot{\Phi}_2 &= \eta_3 \sin(\Phi_2) + \eta_4 \sin(\tfrac{\Phi_1-\Phi_2}{k_0}) \end{aligned} \tag{5}$$

with η_i-s and k_0 negative.

Claim: The system (5) has a locally stable equilibrium at $(0,0)$.

Proof: Firstly, we can easily verify that $\dot{\Phi}_1(0,0) = 0$ and $\dot{\Phi}_2(0,0) = 0$, which means the origin $(0,0)$ is an equilibrium. Then we need to prove the equilibrium $(0,0)$ is stable locally.

Define the Lyapunov function $V(\Phi_1,\Phi_2)$ as follows:

$$V(\Phi_1,\Phi_2) = \frac{\alpha}{2}\Phi_1^2 + \frac{1}{2}\Phi_2^2$$

with $\alpha = \frac{\eta_4}{\eta_2} = \frac{|\eta_4|}{|\eta_2|} > 0$. It is obvious that $V(\Phi_1,\Phi_2)$ is positive definite. Then we have

$$\begin{aligned} \dot{V}(\Phi_1,\Phi_2) &= \alpha\Phi_1\dot{\Phi}_1 + \Phi_2\dot{\Phi}_2 \\ &= \alpha\Phi_1\left(\eta_1 \sin(\Phi_1) + \eta_2 \sin(\tfrac{\Phi_2-\Phi_1}{k_0})\right) \\ &\quad +\Phi_2\left(\eta_3 \sin(\Phi_2) + \eta_4 \sin(\tfrac{\Phi_1-\Phi_2}{k_0})\right) \\ &= \alpha\eta_1\Phi_1 \sin(\Phi_1) + \eta_3\Phi_2 \sin(\Phi_2) \\ &\quad +|\eta_4|(\Phi_1-\Phi_2)\sin(\tfrac{\Phi_2-\Phi_1}{|k_0|}) \end{aligned}$$

with η_i-s and k_0 negative, we can easily verify that

$$\begin{aligned} \alpha\eta_1\Phi_1 \sin(\Phi_1) &< 0 \\ \eta_3\Phi_2 \sin(\Phi_2) &< 0 \\ |\eta_4|(\Phi_1-\Phi_2)\sin(\tfrac{\Phi_2-\Phi_1}{|k_0|}) &< 0 \end{aligned}$$

for (Φ_1,Φ_2) with $|\Phi_1|$ and $|\Phi_2|$ small enough. Then $\dot{V}(\Phi_1,\Phi_2)$ is negative definite for (Φ_1,Φ_2) in the neighborhood of the origin $(0,0)$. By Lyapunov Theorem, the equilibrium $(0,0)$ is locally stable in the system (5). Hence, the 'scaled' phase differences, $k_0k_1\phi_1 - \phi_0$ and $k_0\phi_2 - \phi_1$, converge to $-\psi_1$ and $-\psi_3$ in the system (3).

In the dynamical model of oscillatory network with the Fourier terms in the first equation (2), the 'scaled' phase differences, $k_0k_1\phi_1 - \phi_0 + \psi_1$ and $k_0\phi_2 - \phi_1 + \psi_3$, can not converge to a certain equilibrium, but they can be controlled within certain range by choosing stronger connection parameters, ε_i-s, in the model (2). For simplicity, we choose $\psi_i = 0$ and $\varepsilon_i = \varepsilon$ for all i in the model (2). By increasing the connection coefficients, ε_i, the 'scaled' phase differences, $k_0k_1\phi_1 - \phi_0$ and $k_0\phi_2 - \phi_1$ can be reduced within smaller range. Fig. 17 shows the 'scaled' phase difference patterns $k_0k_1\phi_1 - \phi_0$ (red), $k_0\phi_2 - \phi_1$ (green) and $k_1\phi_1 - \phi_2$ (blue), for three different connection coefficients, $\varepsilon_i = \varepsilon = 0.2, 0.8$ and 1.6 from top to bottom. and maximal phase differences become smaller as the connection coefficients, ε_i, increase, which is shown in Fig. 18.

5. CONCLUSION

We have analyzed the transcriptome data from Cyanothece, a photosynthetic cyanobacteria, for the purpose of discovering genes whose expressions are rhythmically close (co-rhythmic). The rhythmic component of the micro array expression patterns can be reconstructed using two sinusoidal functions with a pair of optimum frequencies. Subsequently we utilized an evolution equation to model the phase activity of KaiC phosphorylation/dephosphorylation cycle. At last we proposed an oscillatory network, using a cascade of three phase oscillators, to model the phase patterns of rhythmic components in the micro array expression data. One of the phase oscillators 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'. Further work on the oscillatory network will be done. The work will include how the master clock controls the phases of the other two oscillator, if one master clock is enough to drive phase patterns of rhythmic components in multiple micro array expression data, and how those parameters in the model are chosen.

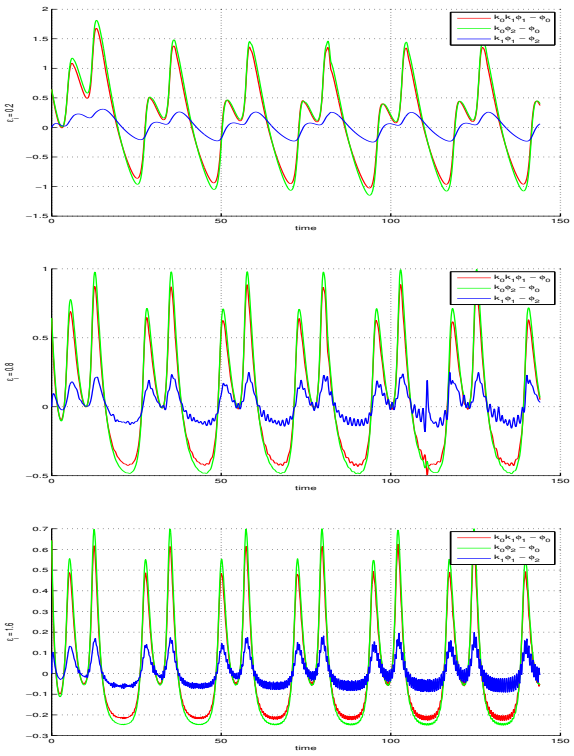

Fig. 17. The 'scaled' phase difference patterns $k_0k_1\phi_1 - \phi_0$ (red), $k_0\phi_2 - \phi_1$ (green) and $k_1\phi_1 - \phi_2$ (blue), for three different connection coefficients, $\varepsilon_i = \varepsilon = 0.2, 0.8$ and 1.6 from top to bottom.

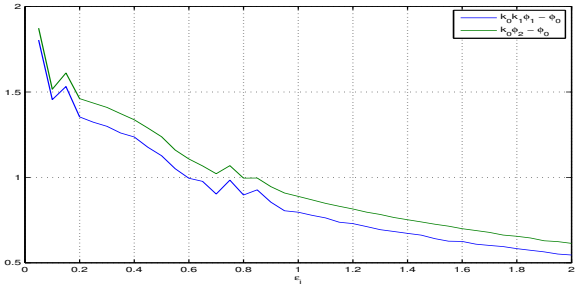

Fig. 18. The maximal phase differences become smaller as the connection coefficients, ε_i, increase

REFERENCES

Y. Kuramoto. *Chemical Oscillations, Waves, and Turbulence*. Springer Verlag, New York, 1984.

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):816–823, 2006.

J. Quackenbush. Microarray data normalization and transformation. *Nature Genetics*, 32:496–501, 2002

P. A. Tass. *Phase Resetting in Medicine and Biology: Stochastic Modelling and Data Analysis*. Springer Verlag, New York, 2006.