Effect of continuous light on diurnal rhythms in *Cyanothece* ATCC 51142

Thanura Elvitigala[1], Jana Stöckel[2], Bijoy K. Ghosh[3] and Himadri B. Pakrasi[2*]

[1]Department of Electrical and Systems Engineering, Washington University, St. Louis, MO 63130, USA;

[2]Department of Biology, Washington University, St. Louis, MO 63130, USA;

[3]Department of Mathematics and Statistics, Texas Tech University, Lubbock, TX 79409.

Email addresses:

TE: tre1@ese.wustl.edu

JS: stockel@biology2.wustl.edu

BKG: bijoy.ghosh@ttu.edu

HBP: pakrasi@wustl.edu

*To whom correspondence should be addressed: Himadri B. Pakrasi, pakrasi@wustl.edu

Pages: 26

Words count: 5273

ABSTRACT

Background: Life on earth is strongly affected by alternating day and night cycles. Accordingly, many organisms have evolved an internal timekeeping system with a period of approximately 24 hours. Cyanobacteria are the only known prokaryotes with robust rhythms under control of a central clock. Numerous studies have been conducted to elucidate components of the circadian clock and to identify circadian-controlled genes. However, the complex interactions between endogenous circadian rhythms and external cues are currently not well understood, and a direct and mathematical based comparison between light-mediated and circadian-controlled gene expression is still outstanding. Therefore, we combined and analyzed data from two independent microarray experiments, previously performed under alternating light-dark and continuous light conditions in *Cyanothece* ATCC 51142, and sought to classify light responsive and circadian controlled genes.

Results: Fourier Score-based methods together with random permutations and False Discovery Rates were used to identify genes with oscillatory expression patterns, and an angular distance based criterion was applied to recognize transient behaviors in gene expression under constant light conditions. Compared to previously reported mathematical approaches, the combination of these methods also facilitated the detection of modified amplitudes and phase-shifts of gene expression. Our analysis showed that the majority of diurnally regulated genes, essentially those genes that are maximally expressed during the middle of the light and dark period, are in fact light responsive. In contrast, most of the circadian controlled genes are up-regulated during the beginning of the dark or subjective dark, and are greatly enriched for genes

associated with energy metabolism. Genes involved in biogenesis of the bidirectional hydrogenase are under circadian control, whereas genes related to the uptake hydrogenase reveal an altered gene expression under constant light. Many of the circadian controlled and light responsive genes are found in gene clusters within the *Cyanothece* ATCC 51142 genome. Interestingly, in addition to cyclic expression patterns with a period of 24 hours, we also found several genes that oscillate with an ultradian period of 12 hours, a novel finding among cyanobacteria.

Conclusions: We demonstrate that a combination of different analytical methods significantly improved the identification of cyclic and transient gene expression in *Cyanothece* ATCC 51142. Our analyses provide an adaptable and novel analytical tool to study gene expression in a variety of organisms with diurnal, circadian and ultradian behavior.

BACKGROUND

Temporal periods of light have provided a significant abiotic selection pressure and greatly influenced evolution on earth. Accordingly, diurnal changes in various biological activities have been observed in different organisms, ranging from bacteria to mammals. Central for the capacity to anticipate such environmental changes is an internal clock that controls circadian gene expression and orchestrates cell physiology in synchrony with a day and night cycle.

Cyanobacteria are the only prokaryotes with a circadian clock (1). These organisms thrive in environments as diverse as fresh and salt waters, glaciers, deserts, and hot springs, where they play a major role in global carbon sequestration and oxygen production. In addition, some cyanobacteria, such as *Cyanothece* sp. ATCC 51142 (*Cyanothece* 51142 hereafter) are able to fix atmospheric nitrogen and contribute to a great extent to the marine nitrogen cycle (2). *Cyanothece* 51142 is a unicellular cyanobacterium that separates oxygenic photosynthesis and nitrogen fixation temporally, thereby circumventing the irreversible inactivation of the oxygen sensitive nitrogenase enzyme (3). In *Cyanothece* 51142, photosynthesis occurs during the day and nitrogen fixation during the night of a diurnal cycle (4).

Previous work has focused on the identification of different clock components such as *kaiA*, *kaiB* and *kaiC*, and several detailed mathematical models describing the interactions between oscillator components have been published. Most of these studies are based on modeling transcriptional and translational feedback loops as well as post-transcriptional modifications of clock genes (5-7). In other work, global DNA microarray analysis of *Synechocystis* sp. PCC 6803, a non-diazotrophic strain, under constant light

4

conditions determined the extent of circadian control at the transcriptional level, leading to the prediction of 2-7% of cyclic genes (8). In contrast, microarray analyses of *Cyanothece* 51142 under alternating light and dark conditions revealed that ~30% of the genes in the genome are significantly diurnally regulated (9), whereas only 10% of these genes are circadian controlled (10). However, a detailed analytical approach to distinguish between diurnal and circadian regulated genes in *Cyanothece* 51142 is still outstanding.

Organisms such as *Cyanothece* 51142 perform different cellular processes at distinct phases of the diurnal cycle and thus depend on tightly regulated expression programs. On the other hand, the capability to adjust to periodic changes in their environment, in particular to alterations of light intensity and quantity, provides an important basis for the ecological success of these organisms. Light is an essential abiotic factor, not only because cyanobacteria obtain their entire energy through photosynthesis, but also for the activation and regulation of many central metabolic processes. In order to understand how *Cyanothece* 51142 balances gene expression under nitrogen-fixing conditions in consecutive light and dark cycles with altered physiological requirements under constant light conditions, we analyzed data sets from two different microarray experiments, to elucidate details beyond of what has been found in the former studies (9-11).

For this study we developed various advanced mathematical tools to identify rhythmic and transient patterns of gene expression. We demonstrate that a combination of different analytical approaches substantially improves the distinction between transient and cyclic gene expression. Our analysis revealed that the expression of 63% of the

previously observed strongly diurnally regulated genes (9) is significantly affected under constant light conditions. Interestingly, the majority of circadian controlled genes in *Cyanothece* 51142 are maximally expressed during the subjective dark and largely comprised of genes related to energy metabolism. Furthermore, we found a significant number of genes that show ultradian rhythms in their expression and oscillate with a period of 12 hours. This study provides a novel and adaptable analytical tool for comparison of data from different microarray experiments to study gene expression in a variety of organisms with diurnal, circadian and ultradian behavior.

RESULTS AND DISCUSSION

Microarray data sets.   We have analyzed two independent *Cyanothece* 51142 microarray data sets that were performed over two consecutive diurnal periods with a sampling rate of every four hours and a shift in sampling time of two hours between the experiments.   The studies were conducted using Agilent (www.agilent.com) custom-made two-channel microarrays.   The work of Stöckel *et al.* (9) focused on global gene expression in *Cyanothece* ATCC 51142 under alternating 12 hour light and dark cycles, while Toepel *et al.* (10) investigated changes in gene expression under a 12 hour light and 12 hour dark period followed by a constant light period of 24 hours.   In both experiments *Cyanothece* 51142 cells were grown under nitrogen-fixing conditions.

Data analyses and identification of cyclic expressed genes. In Stöckel *et al.* (9) diurnal regulated genes were classified using a correlation network and a fold-change cutoff of 1.3, in which genes with a period other than 24 hours were excluded, since they did not correlate with the main oscillatory network. In Toepel *et al.* (10, 11) cyclic genes were identified using a 2-fold cutoff for maximal changes in gene expression. Although the main oscillatory behaviors are detectable using these methods, alterations in gene expression such as changes of amplitudes and phase-shifts were not observed, especially because the altered light conditions were applied only for a short period of time. Therefore we proposed a combination of an angular distance based criteria to characterize such alterations in gene expression.

After normalization of both raw data sets, the main frequency components of gene expression were identified using Fast Fourier Transform (FFT) according to (12). For

the data from Stöckel *et al.* (9), the majority of genes under alternating light and dark cycles oscillate with a period of approximately 24 hours (Figure 1a). Interestingly, a considerable number of genes show a principal frequency close to two cycles per day, which corresponds to genes that oscillate with a period of 12 hours. In comparison, similar analyses of data from Toepel *et al.* (10) show that the frequencies are distributed over a wider range, with a larger number of genes cycling with 12 and 24 hour periods (Figure 1b). The extended periods of 40-48 hours (frequencies between 0.5 and 1 cycles per day) are mainly assigned to genes that do not continue to oscillate under continuous light conditions. Thus, the differences between the two histograms primarily reflect the impact of continuous light on gene expression in *Cyanothece* 51142.

The two main frequencies of 12 and 24 hours were used for further Fourier score calculations. For the determination of cyclic behavior, a Fourier score cutoff of 6.5 corresponding to a False Discovery Rate of 2% under alternating light and dark conditions and of 3% under continuous light was used. Based on this analysis, 659 out of a total of 4888 genes that oscillate with a period of 24 hours and 50 genes that cycle with a period of 12 hours were identified as common in both microarray datasets (Supplemental Tables 1 and 2).

Classification of light responsive genes. Since the Fourier score-based method is primarily designed for the identification of cyclic gene expression, and because genes with altered expression patterns are detectible only in time points corresponding to the subjective dark, an angular distance based method was proposed to identify transient behaviors in gene expression. Accordingly, the data were separated into four 12 hour subsets, which correspond to the individual light and dark periods and thus resulted in

four 3-dimensional vectors. If a gene shows a cyclic expression, smaller distances between vectors related either to two light or two dark periods are expected. In contrast, larger distances are anticipated for vectors from different periods such as light and dark. For genes showing transient behavior in their expression, the vector equivalent to the subjective dark is likely to be different from the dark (Table 1). Figure 2a shows the distribution of vectors for the gene *cce_2318*, which encodes HoxF, one of the subunits of the bidirectional hydrogenase, as an example of circadian gene expression. All vectors corresponding to individual light and dark periods are well separated and the vector representing the subjective dark is closely associated with the vectors derived from other dark periods. In contrast, the gene *cce_1063,* encoding the small subunit HupS of the uptake hydrogenase, does not continue to oscillate under constant light conditions (Figure 2b). The altered expression of *cce_1063* is clearly indicated in that the vector related to the subjective dark period shows a different orientation.

The similarity threshold chosen for the angular distance method ought to yield a dataset of cycling genes in utmost agreement between the Fourier score based and angular distance methods. Therefore, for the angular distance method the number of cyclic genes among the set of already identified cycling genes was calculated for different threshold values (Figure 3). Based on these calculations, a cutoff of 0.8 for the angular distance was selected, which resulted in 97% and 78% Fourier score based agreement for the Stöckel *et al.* (9) and Toepel *et al.* (10) data sets, respectively. The expression of a gene in two different 12 hour periods was considered to be similar if the corresponding vectors were within a distance of 0.8 to each other, and disparate if the

distance was higher. The vectors of genes transcribed with an ultradian period of 12 hours revealed consistently smaller angular distances for any two regions (Table 1).

Based on the Fourier score and angular distance based methods, four major groups of genes could be identified. The genes in group 1 continued to oscillate under continuous light and were classified as circadian controlled genes, whereas genes from group 2 were light responsive and showed oscillating expression only under alternating light and dark cycles. Both groups represent genes that cycle with a 24 hour period. In contrast, genes from group 3 and 4 oscillate with an ultradian period of about 12 hours, with genes from group 3 being circadian controlled and the genes from group 4 light responsive (Table 1, Supplemental Table 2).

Our analysis of the Stöckel *et al.* dataset (9) identified 43% of genes in the *Cyanothece* 51142 genome (2138 genes) with oscillating expression patterns under alternating light and dark conditions, in comparison to the previously reported 1445 genes. This suggests that diurnal regulation of gene expression in *Cyanothece* 51142 might be greater than previously thought. However, after combining and analyzing both data sets using the two different methods, only 772 (14.3%) of genes in the genome were found to be diurnally regulated or light inducible, and 448 genes (9%) could be classified as circadian controlled. This relatively small number of diurnally regulated genes common in both data sets could result from differences in growth and culture conditions and comprises therefore the more significant cyclic expressed genes. Interestingly, five circadian controlled genes and 45 genes with transient expression patterns oscillate with an ultradian frequency of 12 hours (Supplemental Table 2). As an example, figure

4 shows the expression pattern for the genes *cce_3226* and *cce_1889* that oscillate with a 12 hour period.

Taken together, the combination of the angular distance and Fourier Score based methods lent an additional level of confidence to the identification of cyclic expressed genes in *Cyanothece* 51142. These analyses uncovered that most of the previously identified diurnal genes are indeed light responsive.

Genes belonging to different functional categories. Considering the central role of photosynthesis in the cellular metabolism of *Cyanothece* 51142 and other cyanobacteria, a great impact of light on gene expression was anticipated. We found that the majority of regulatory genes are light inducible, presumably the result of comprehensive changes that take place in order to adjust to altered light conditions (Supplemental Table 1). In addition, the expression of many genes related to transcription and translation is light regulated and likely accounts for a higher turnover of proteins under continuous light. Our analyses also revealed that several genes associated with photosynthesis are circadian controlled. For instance, *petC1*, one of the genes encoding the iron-sulfur subunit of the cytochrome b_6f complex, is circadian controlled, whereas *petC2* was classified as light responsive (Supplemental Table 1). Similarly, a number of ATP-synthase subunits, including *atpF* and *atpH,* are encoded by multicopy genes, with one gene copy being diurnal regulated while the other gene copy was found to be circadian controlled (Supplemental Table 1). The gene *atpH* has been identified as clock controlled in previous studies of *Synechocystis* PCC 6803 (8).

A more detailed analysis of different metabolic processes revealed that nitrate and sulfate assimilation are largely under circadian control (Supplemental Table 1).

Furthermore, the genes *hoxEFUYH* encoding the bidirectional hydrogenase show a significant cyclic behavior in their expression (Supplemental Figure 1a, b). In comparison, a recent microarray analysis performed in *Cyanothece* 51142 over 24 hours under alternating 6 hours light/dark conditions led the authors to the conclusion (11) that the expression of the *hox* genes under their conditions are not circadian. However, an alternative interpretation of their data (Figure 3B in (11)) suggests that the *hoxE* gene does in fact show an oscillatory expression with a period of 24 hours, which would sustain our findings. This is also supported by previous studies in *Synechococcus* PCC 7942 which uncovered a strong circadian component in the expression of *hoxF* and *hoxE* (13). In total, our study revealed that 9 out of 17 genes involved in hydrogen metabolism in *Cyanothece* 51142 were expressed at the same level during the dark and subjective dark and therefore classified as circadian controlled. In contrast, the expression of *hupS* and *hupL* encoding subunits of the uptake hydrogenase are severely affected during the subjective dark, with *hupS* showing an 8-time lower expression level (Supplementary Figure 1d). This is notable, since genes encoding the uptake hydrogenase and genes involved in nitrogen fixation are highly co-expressed under alternating light and dark conditions (9) and several studies point to close functional interactions of both enzymes (14). However, earlier studies of different *Nostoc* species uncovered that the expression of genes encoding the uptake hydrogenase is substantially stimulated by supplemented hydrogen (15). Accordingly, the reduced transcript abundance of *hupS* and *hupL* is presumably coupled to a decline in the activity of the nitrogenase enzyme, which has been observed under continuous light (10). Thus, it becomes evident that the expression of genes

associated with nitrogen fixation and genes encoding the uptake hydrogenase are regulated differentially under certain growth conditions. This is also supported by the fact that *hupS* in *Cyanothece* 51142 is expressed under both nitrogen sufficient and deficient conditions, whereas the expression of nitrogenase related genes such as *nifH* is not (data not shown). Furthermore, even though the genes involved in nitrogen fixation are organized in a single cluster consisting of two adjacent regulons on opposite strands in the genome (16), different expression profiles for many of those genes were observed under continuous light. The genes *nifB*, *nifS*, *nifE*, *nifN*, *nifX* and *nifW,* which are involved in biogenesis and assembly of the Mo-Fe cofactor, revealed a more than two fold increase in their expression levels (Toepel et al., 2008) in addition to a shift in phase of about four hours under continuous light conditions (Supplemental Figure 2d).

In fact, several light responsive and circadian controlled genes are found in clusters of three or more genes throughout the *Cyanothece* 51142 genome (Supplemental Figures 3, 4; Supplemental Table 1). Such clustering of similarly regulated genes into different regions of the genome provides an opportunity for the cell to express genes more efficiently. On the other hand, additional diurnal and circadian controlled genes can be identified which might not have been detected due to low amplitudes of oscillations. Thus, there is a high probability that an unclassified gene within a group of genes assigned to one of the two categories belongs to the same class.

Furthermore, we found that the majority of circadian controlled genes are maximally expressed at the beginning of the dark period and are largely comprised of genes related to energy metabolism (Figure 5a). In contrast, most of the light responsive genes peak during the middle of the dark or light periods (Figure 5b), which suggests

that their expression is regulated by the presence or absence of light rather than in anticipation of the different light cycles. In addition, FFT, Fourier Score, and angular distance calculations revealed 50 genes that show ultradian rhythms in their expression and oscillate with a period of 12 hours. Ultradian rhythms of cell division have been observed in several organisms including cyanobacteria (17), even though the major timing system for all light-sensitive organisms is a circadian clock that operates with a period of approximately 24 hours. Therefore, the finding of 12 hour cycles in the expression of genes related to other cellular processes was unanticipated and is novel among cyanobacteria. Previous studies in eukaryotic unicellular microbes such as yeast have shown that an ultradian oscillator regulates cyclic respiratory activity and global gene expression (18, 19). This is particularly interesting since *cydB*, which encodes a subunit of one of the terminal oxidases in *Cyanothece* 51142, shows an ultradian expression in both microarrays datasets (Figure 4b), whereas the expression of the majority of respiratory genes is considerably affected under subjective dark conditions (10). However, further studies are required to elucidate the role and physiological significance of such ultradian rhythms in *Cyanothece* 51142.

CONCLUSIONS

In this work we compared two independent microarray experiments to study changes in gene expression under alternating and constant light conditions in *Cyanothece* 51142. Two mathematical approaches were applied to differentiate transient behaviors from cyclic gene expression. Fourier score based methods can successfully be used for identification of cyclic behaviors but they fail to detect some transient behaviors such as shifts of the expression peaks or smaller amplitudes under constant light conditions. Therefore, an angular distance based method was applied which improved the detection of such changes. These methods are valid for distinguishing transient from cyclic behaviors even for the limited number of time points in one of the different light treatments. This approach provides an adaptable tool for future studies of diurnal, circadian and ultradian rhythms in various organisms. Furthermore, the combined analysis of two different microarray data sets enabled the extraction of novel details about diurnal and circadian regulation in *Cyanothece* 51142 which would not be obtained from the individual experiments alone.

FIGURE LEGENDS

Figure 1: Distribution of principal frequencies. Fast Fourier Transform was used to determine the principal frequencies of different gene expressions. (a) Histogram of principal frequencies obtained from the microarray data set under alternating light and dark cycles (9). (b) Distribution of principal frequencies from the Toepel *et al.* data set (10) under alternating light and dark periods followed by continuous light. The x-axis shows the frequency and the y-axis shows the number of genes.

Figure 2: Distribution of vectors corresponding to different light and dark periods for two distinct hydrogenase genes. (a) The gene *cce_2318* shows a cyclic expression profile under alternating light and dark conditions and does not reveal significant changes during the subjective dark period. (b) The expression of the gene *cce_1063* alters considerably under subjective dark. *Red* colored vectors represent the different light periods, *Blue* colored vectors correspond to the dark periods, and *Green* vectors indicate subjective dark.

Figure 3: Threshold evaluation. The number of cyclic expressed genes using different cutoffs for the angular distance method are shown. The threshold value resulting in the maximum number of oscillating genes for both microarray data sets was chosen for the final classification. The blue graph corresponds to the cyclic genes from the Stöckel *et al.* data set (9) and the green graph shows the oscillating genes from the Toepel *et al.* data set (10).

Figure 4:   Ultradian cycles in gene expression.   The genes (a) *cce_3226* and (b) *cce_1889* reveal ultradian rhythms in their gene expression and oscillate with a period of 12 hours.

Figure 5:   Peak time distribution for circadian controlled and light responding genes.   The graphs represent the peak time data for genes which were classified as (a) circadian controlled and (b) light responsive.   The circle defines a polar coordinate plot with radians transformed to hours.   The inner dotted circles represent a scale for the number of genes.   The time points are plotted around the circle. Most of the circadian genes peak in the early dark, whereas light responding genes peak in the middle of either light or dark.

Table 1. Average distance from different light and dark periods in each of the four classified gene groups.

Group	Stöckel *et al.* data set (9)						Toepel *et al.* data set (10)					
	D_1L_1	D_1D_2	D_1L_2	L_1D_2	L_1L_2	D_2L_2	L_1D_1	L_1L_2	L_1SD_2	D_1L_2	D_1SD_2	L_2SD_2
1	1.78	0.11	1.79	1.85	0.08	1.79	1.79	0.12	1.68	1.70	0.33	1.69
2	1.78	0.17	1.72	1.88	0.14	1.76	1.73	0.18	1.57	0.68	1.42	0.78
3	0.23	0.44	0.29	0.37	0.05	0.44	0.58	0.17	0.27	0.46	0.68	0.27
4	0.33	0.17	0.29	0.43	0.20	0.30	1.44	0.40	0.63	1.26	1.04	0.92

D: dark; L: light; SD: subjective dark

METHODS

Experimental datasets

For the Stöckel *et al.* microarray data set (9), *Cyanothece* 51142 cultures were grown in ASP2 medium without NaNO$_3$ (3) at 30°C with air bubbling in alternating 12 hour light and dark cycles with 50 μmol photons/m^2*s of white light. After 7 days, 150 mL samples were collected every 4 hours for 2 days, starting with 1 hour into the dark period (time point D1). In total, 12 samples were collected.

For the Toepel *et al.* data set (10), *Cyanothece* 51142 cultures were grown in an airlift bioreactor (6L BioFlo 3000; New Brunswick Scientific, Edison, NJ) in ASP2 medium without NO$_3$ at 30°C in 12 hour light and 12 hour dark cycles. The culture was illuminated by two light-emitting-diode panels using orange (640 nm) and red (680 nm) light, yielding an intensity of 100 μmol photons/m^2*s inside the bioreactor. The cultures were grown for 5-6 days under alternating light and dark conditions prior to collecting samples for every 4 hours over a time period of 2 days, starting at the transition from dark to light (time point L0) for 24 hours followed by 24 hours of continuous light.

Algorithms and data processing

Preliminary Data Processing. In previous studies (9-11), the data were processed and analyzed differently, which made a direct comparison of the results difficult. Therefore, in this communication both data sets have been combined and analyzed uniformly. The raw data were normalized using the LOWESS normalization algorithm to avoid a systematic intensity based bias which is commonly observed with two-channel

microarrays (20) and preliminary data processing of both data sets was performed according to Stöckel et al. (2008).

Identification of Cyclic Genes using Fourier Score

In previous studies (21), several cycle detection methods were compared and the Fourier score based method with random permutations was shown to be more appropriate for situations where data were available only for a few cycles. The Fourier score for any given time course signal x_i at a given frequency f was defined as:

$$F_i = \sqrt{\left(\sum_t{}' \sin \omega t \cdot x_i(t)\right)^2 + \left(\sum_t{}' \cos \omega t \cdot x_i(t)\right)^2}$$

where $\omega = 2\pi f$.

Initial Fast Fourier Transform calculations indicated existence of oscillations with 24h and 12h periods. Therefore Fourier score calculations were performed using those frequencies.

If a given signal consists of a dominant cyclic component of the corresponding frequency, a larger Fourier score is expected. The significance of the Fourier score was quantified by comparing the Fourier score of the original signal with scores of large collection of random signals. These random signals were obtained by using different permutations of the original signal. The original gene expressions were scaled to have a unit standard deviation, which enabled direct comparisons of Fourier scores from different genes.

20

A False discovery rate (FDR) based approach similar to (22) was used to determine a global threshold for the Fourier scores, in order to separate cyclically expressed genes from transiently expressed genes. An empirical FDR for the Fourier score at a chosen threshold *t* was defined as:

$$FDR(t) = \frac{\sum_{k=1}^{N}\sum_{j=1}^{M} I(F_{j,k} \geq t)/M}{\sum_{k=1}^{N} I(F_k^0 \geq t)}$$

.

Hereby corresponds *M* to the number of permutations which were used for the null hypothesis. M=10,000 were chosen for all calculations. *N* represents the total number of genes, *F* the Fourier score for a random signal *j* which was obtained for a gene *k*, and *Fj* the Fourier score of the original expression signal for gene *k*. *I(x)* represents an indicator function as follows:

$$I(x) = \begin{cases} 1 & if\ x > 0 \\ 0 & otherwise \end{cases}$$

.

Identification of transient behaviors using angular distance. We propose an angular distance based method to classify the transient gene expression under constant light conditions. The data were separated into four 12 hour data sets, which correspond to the different light and dark periods and resulted in four 3-dimensional vectors for each gene and for each experiment. The pair wise angular distances between different vectors for a given gene were calculated as:

$$d_{1,2} = \left(1 \quad x_1 x_2^T/(x_1 x_1^T)^{1/2}(x_2 x_2^T)^{1/2}\right)$$

,

where x_1 and x_2 are the vectors from two different 12 hour periods. $D_{1,2}$ can have any value between 0 and 2, with 0 representing vectors with the same direction and 2 representing vectors with the opposite direction.

AUTHOR'S CONTRIBUTION

T.E., J.S., H.B.P. and B.K.G. designed research; T.E. and J.S. performed data analysis; T.E. and J.S. wrote the paper. All authors read and approved the final manuscript.

REFERENCES

1. Kondo T, Strayer CA, Kulkarni RD, Taylor W, Ishiura M, Golden SS, Johnson CH: **Circadian rhythms in prokaryotes: luciferase as a reporter of circadian gene expression in cyanobacteria.** *Proc Natl Acad Sci USA* 1993, **90:**5672-5676.

2. Zehr JP, Waterbury JB, Turner PJ, Montoya JP, Omoregie E, Steward GF, Hansen A, Karl DM: **Unicellular cyanobacteria fix N2 in the subtropical North Pacific Ocean.** *Nature* 2001, **412:**635–638.

3. Reddy KJ, Haskell JB, Sherman DM, Sherman LA: **Unicellular, aerobic nitrogenfixing cyanobacteria of the genus *Cyanothece*.** *J Bacteriol* 1993, **175:**1284–1292.

4. Sherman LA, Meunier P, Colon-Lopez MS: **Diurnal rhythms in metabolism: A day in the life of a unicellular, diazotrophic cyanobacterium.** *Photosynth Res* 1998, **58:** 25-42.

5. Mehra A, Hong CI, Shi M, Loros JJ, Dunlap JC, Ruoff P: **Circadian rhythmicity by autocatalysis.** *PLoS Computational Biology* 2006, **2:**e96.

6. Miyoshi F, Nakayama Y, Kaizu K, Iwasaki H, Tomita M: **A Mathematical Model for the Kai-Protein-Based Chemical Oscillator and Clock Gene Expression Rhythms in Cyanobacteria.** *J Biol Rhythms* 2007, **22:**69-80.

7. Rust MJ, Markson JS, Lane WS, Fisher DS, O'Shea EK: **Ordered Phosphorylation Governs Oscillation of a Three-Protein Circadian Clock.** *Science* 2007, **318:**809-812.

8. Kucho K, Okamoto K, Tsuchiya Y, Nomura S, Nango M, Kanehisa M, Ishiura M: **Global analysis of circadian expression in the cyanobacterium *Synechocystis* sp. strain PCC 6803.** *J Bacteriol* 2005, **187:**2190-2199.

9. Stöckel J, Welsh EA, Liberton M, Kunnvakkam R, Aurora R, Pakrasi HB: **Global transcriptomic analysis of Cyanothece 51142 reveals robust diurnal oscillation of central metabolic processes.** *Proc Natl Acad Sci USA* 2008, **105:**6156-6161.

10. Toepel J, Welsh E, Summerfield TC, Pakrasi HB, Sherman LA: **Differential transcriptional analysis of the cyanobacterium *Cyanothece* sp. strain ATCC 51142 during light-dark and continuous-light growth.** *J Bacteriol* 2008a, **190:**3904-3913.

11. Toepel J, McDermott J, Summerfield TC, Sherman LA: **Transcriptional analysis of the unicellular diazotrophic cyanobacterium *Cyanothece* sp. ATCC 51142 grown under short day/night cycles.** *J Phycology* 2008b, *in press.*

12. Duhamel P, Vetterli M: **Fast Fourier Transforms: A Tutorial Review and a State of the Art.** *Signal Processing* 1990, **19:**259-299.

13. Schmitz O, Boison G, Bothe H: **Quantitative analysis of expression of two circadian clock-controlled gene clusters coding for the bidirectional hydrogenase in the cyanobacterium *Synechococcus* sp. PCC7942.** *Mol Microbiol* 2001, **41:**1409-1417.

14. Tamagnini P, Leitao E, Oliveira P, Ferreira D, Pi*nto F,* Harris DJ, Heidorn T, Lindblad P: **Cyanobacterial hydrog*enases: diversity, reg*ulation and applications.** *FEMS Microbiol Rev* 2007, **31:**692-720.

15. Axelsson R, Lindblad P: **Transcriptional regulation of *Nostoc* hydrogenases: effects of oxygen, hydrogen, and n*ickel. Appl* Environ Microbiol* 2002, **68:**444-447.

16. Welsh EA, Liberton M, Stöckel J, Loh T, Elvitigala T, Wang C, Wollam A, Fulton RS, Clifton SW, Jacobs JM, Aurora R, Ghosh BK, Sherman LA, Smith RD, Wilson RK, Pakrasi HB: **The genome of *Cyanothece* 51142, a unicellular diazotrophic cyanobacterium important in the marine nitrogen cycle.** *Proc Natl Acad Sci USA* 2008, **105:**15094-15099.

17. Lloyd D: **Circadian and ultradian clock-controlled rhythms in unicellular microorganisms.** *Adv Microb Physiol* 1998, **39:**291-338.

18. Kippert F, Hunt P: **Ultradian clocks in eukaryotic microbes: from behavioural observation to functional genomics.** *Bioessays* 2000, **22:**16-22.

19. Schibler U, Naef F: **Cellular oscillators: rhythmic gene expression and metabolism.** *Curr Opin Cell Biol* 2005 **17:**223-229.

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

21. de Lichtenberg U, Jensen J, Fausbøll A, Jensen TS, Bork P, Brunak S: **Comparison of computational methods for the identification of cell cycle-regulated genes.** *Bioinformatics* 2005, **21:**1164–1171.

22. Benjamini Y, Hochberg Y: **Controlling the false discovery rate: a practical and powerful approach to multiple testing.** *J Roy Stat Soc B* 1995, **57:**289–300.
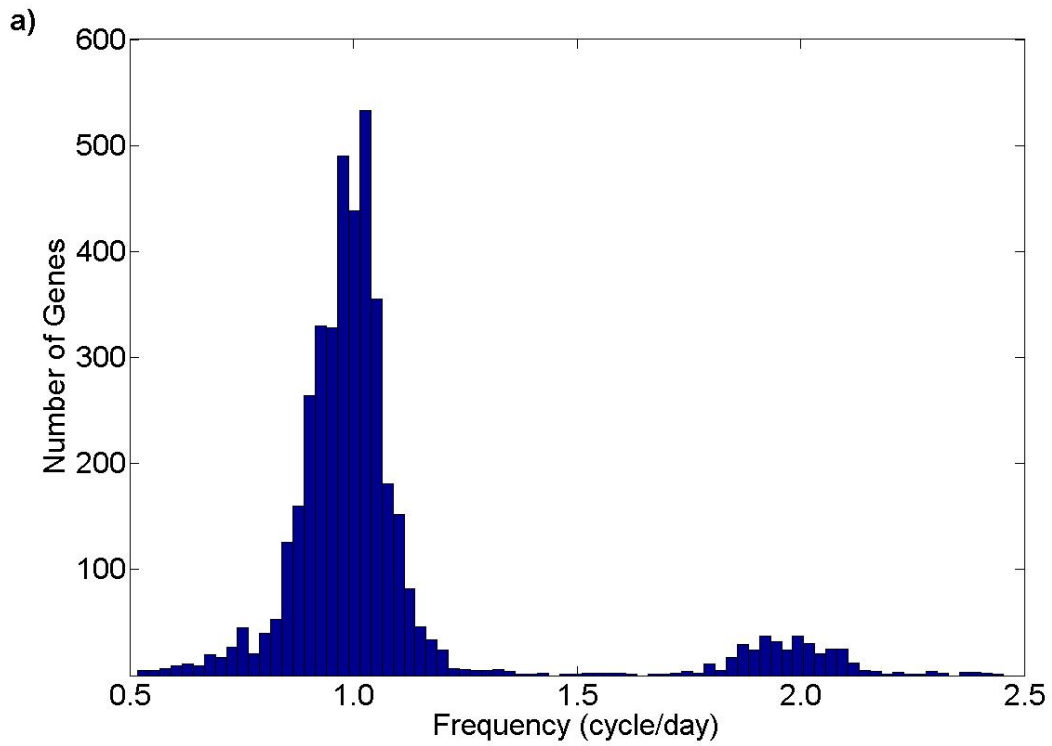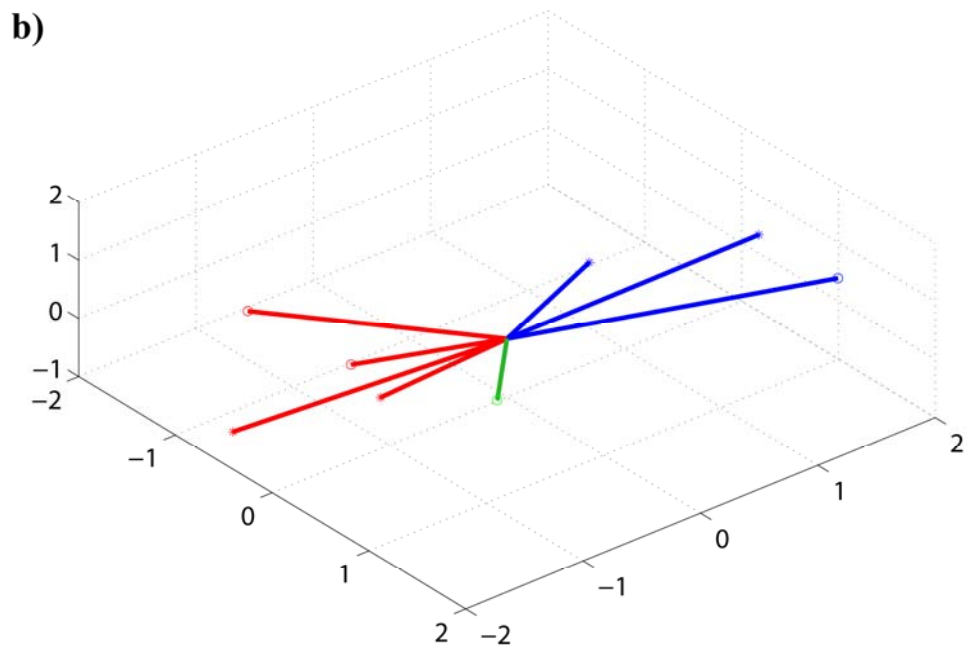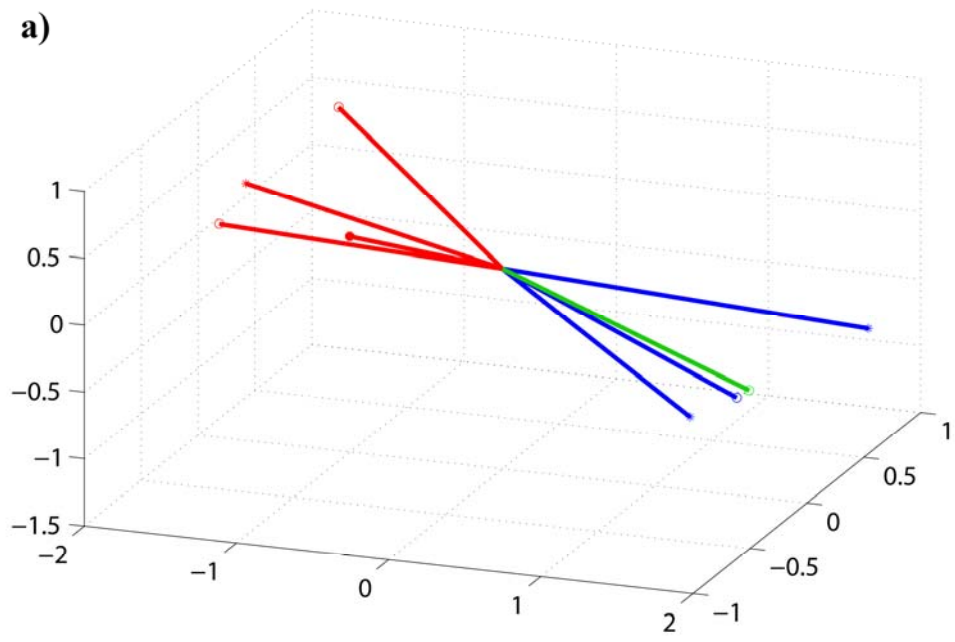
Figures:

Figure 1

a)

b)
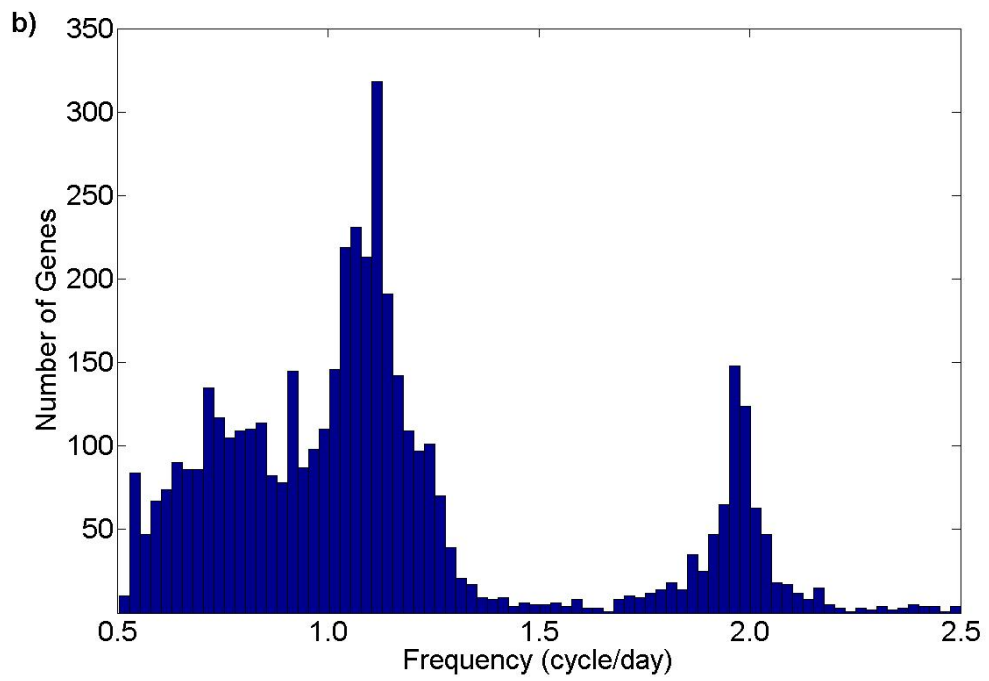
Figure 2

Figure 3

Figure 4

Figure 5