Minimal Tool Set for a Prokaryotic Circadian Clock

Nicolas M Schmelling[1], Anika Wiegard[1], Robert Lehmann[2], Paushali Chaudhury[3], Christian Beck[2], Sonja V Albers[3], and Ilka M Axmann*[1]

[1]Institute for Synthetic Microbiology, Cluster of Excellence on Plant Sciences (CEPLAS), Heinrich-Heine-University Düsseldorf, Universitätsstr. 1, 40225 Düsseldorf, Germany

[2]Institute for Theoretical Biology, Humboldt University Berlin, Invalidenstr. 43, 10115, Berlin, Germany

[3]Molecular Biology of Archaea, University of Freiburg, Institute of Biology II, Schänzlestr.1, 79104 Freiburg, Germany

September 13, 2016

Abstract

Circadian clocks can be found in almost all organisms including photosynthetic Cyanobacteria, whereby large diversity exists within the protein components involved. In the model cyanobacterium *Synechococcus elongatus* PCC 7942 circadian rhythms are driven by a unique KaiABC protein clock, which is embedded in a network of input and output factors. Homologous proteins to the *Synechococcus elongatus* PCC 7942 clock components have been observed in Bacteria and Archaea, where evidence for circadian behavior in these domains is accumulating. However, interaction and function of non-cyanobacterial Kai-proteins as well as homologous input and output components remain mainly unclear.

Using a universal BLAST analyses, we identified putative KaiC-based timing systems in organisms outside as well as variations within Cyanobacteria. A systematic analyses of publicly available microarray data elucidated interesting variations in circadian gene expression between different cyanobacterial strains, which might be correlated to the diversity of genome encoded clock components. Based on statistical analyses of co-occurrences of the clock component homologous to *Synechococcus elongatus* PCC 7942, we propose putative networks of reduced and fully functional clock systems. Further, we studied KaiC sequence conservation to determine functionally important regions of diverged KaiC homologs. Biochemical characterization of exemplary cyanobacterial KaiC proteins as well as homologs from two thermophilic Archaea demonstrated that kinase activity is always present. However, a KaiA-mediated phosphorylation is only detectable in KaiC1 orthologs.

Our analysis of 11,264 genomes clearly demonstrates that components of the *Synechococcus elongatus* PCC 7942 circadian clock are present in Bacteria and Archaea. However, all components are less abundant in other organisms than Cyanobacteria and KaiA, Pex, LdpA, and CdpA are only present in the latter. Thus, only reduced KaiBC-based or even simpler, solely KaiC-based timing systems might exist outside of the cyanobacterial phylum, which might be capable of driving diurnal oscillations.

Introduction

Life on Earth is under the influence of changing environmental conditions, which not only pose a challenge to organisms, but also present a chance of adaptation and therefore a possible fitness advantage over competitors [1, 2]. Using inner timing systems organisms can coordinate their physiology and behavior according to the daily recurring changes. Simple timing systems work in an hour glass like fashion and need to be reset every day by environmental stimuli, whereas true circadian clocks generate self-sustained and temperature-compensated 24-hour rhythms of biological activities [3, 4].

Circadian clocks are found in many eukaryotes such as algae, plants and mammals [5]. Even though circadian clocks seem like a conserved trait in evolution, differences in the protein components, involved in circadian timing, suggest a convergent evolution of timing mechanisms [5]. For many years it has been

*Ilka.Axmann@hhu.de

believed that something as complex as a circadian clock could not have been evolved in unicellular organisms like prokaryotes [5, 6, 7]. However, the existence of temperature compensated 24-hour rhythms of cell division in *Synechococcus* sp. WH 7803 and circadian nitrogen fixation in *Cyanothece* sp. PCC 8801 proved otherwise [8, 9, 10, 11, 12]. The molecular basis of the cyanobacterial circadian clock was intensively investigated in *Synechococcus elongatus* PCC 7942 (hereafter *Synechococcus* 7942), where the core clockwork resembles a posttranslational oscillator [13, 14, 15]. In contrast, eukaryotic circadian rhythms are believed to be mainly based on transcriptional-translational feedback loops. However, findings on post-translational systems are accumulating [16, 17, 18] and might exist also in Archaea [16].

Light is assumed to be the driving stimulus in circadian clock entrainment [19]. In *Synechococcus* 7942, contrary to eukaryotic circadian clock systems, a photoreceptor in the input pathway of the clock could not be detected thus far. Instead, *Synechococcus* 7942 cells sense light indirectly through the redox and energy state of the cell [20]. Here, two metabolic components are considered to play a major role [21]: The ATP to ADP ratio and the redox state of the plastoquinone (PQ) pool [22, 23]. The core of the circadian clock in *Synechococcus* 7942 consists of three proteins KaiA, KaiB and KaiC. KaiC monomers are composed of two domains, which assemble into two hexameric rings [24, 25]. The C-terminal ring is capable of autophosphorylation and –dephosphorylation [26, 27]. KaiC phosphorylation is stimulated by the interaction with KaiA [28, 29], and additionally, affected by the ATP/ADP ratio of the cell [30]. KaiB inhibits the activating effect of KaiA and initializes dephosphorylation [31]. Altogether, KaiC hexamers phosphorylate and dephosphorylate rhythmically during the course of a day. The binding of oxidized quinones to KaiA has been suggested to stop the clock directly by causing KaiA aggregation [32, 33]. The KaiABC core clock is embedded into a network of input and output pathways. The input factors that interact with the core clock are Pex, LdpA, PrkE, NhtA, IrcA, CdpA [20, 34, 35, 36, 37, 38]. Output factors are SasA, LabA, LalA, CpmA, Crm, RpaA, and RpaB, as well as CikA, which is functioning both in input and output pathway of the circadian clock [39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51].

Sequence analysis indicated that at least three different types of timing systems are present in Cyanobacteria, (i) a KaiABC-based system as in *Synechococcus* 7942, (ii) a reduced system with a KaiBC core and a reduced set of input/output factors as in *Prochlorococcus* and (iii) a reduced KaiABC system as in *Synechococcus* sp. WH 8102, which despite including all three *kai* genes, has the same input/output factors as the reduced KaiBC system [52]. Furthermore, multiple *kai* genes can exist in an organism [53, 54]. In *Synechocystis* sp. PCC 6803 (hereafter *Synechocystis* 6803) for example, three copies of both *kaiB* and *kaiC* are found. Besides the core clock components, KaiA, KaiB1 and KaiC1, most similar to the *kai* genes of *Synechococcus* 7942 [55, 56], KaiB3 and KaiC3 are thought to function as fine tuning factors for the circadian clock in *Synechocystis* 6803, whereas no circadian function has been found for KaiB2 and KaiC2 [56]. However, recently *kaiC2*-dependent adaptive growth and diurnal rhythms of nitrogen fixation were observed in *Rhodopseudomonas palustris* [57]. Homologs of *kaiB* and *kaiC* genes also exist in other Bacteria and even Archaea, where a shortend KaiC most often resembles only one domain [53]. An archaeal one-domain KaiC homolog was shown to form a hexameric ring, similar to the two duplicated domains of *Synechococcus* 7942 KaiC [58]. In *Haloferax volcanii* the transcripts of four *kaiC* homologs display diurnal accumulation profiles [59]. However, the function of non-cyanobacterial Kai-proteins is mainly unclear, so far. Although some of the input- and output factors were also found in prokaryotes other than cyanobacteria [49, 52, 60, 61], it is unknown whether the Kai homologs outside the cyanobacterial phylum or the additional cyanobacterial Kai homologs interact with (other) in- and output factors.

In this study, we performed BLAST analyses to first identify possible KaiC-based timing systems in organisms outside of Cyanobacteria and second to explore variations in circadian clocks within Cyanobacteria. Further, we examined variations in circadian gene expression between different cyanobacterial strains using microarray data. Together, this aims at decoding the correlation between Kai proteins and additional clock components. Based on the co-occurrence of clock components known from *Synechococcus* 7942 we propose putative networks of reduced and fully functional clock systems. Further, we used the sequence information of KaiC and its homologs in Cyanobacteria to determine the similarities at important sites of the protein. We chose cyanobacterial KaiC proteins as well as homologs from two thermophilic Archaea and demonstrated that kinase activity is always present. However, a KaiA-mediated phosphorylation is only detectable in cyanobacterial KaiC1 homologs.

Materials and Methods

Programming languages

The programming languages Python (version 3.5.1) and R (version 3.2.3) were used in this work. The processing and analysis of the microarray time series datasets was performed using R. Regarding the distribution analysis, the Biopython project ([62]; version 1.66) was used to download from GenBank as well as to work with FASTA files. Besides Biopython, the Python packages: IPython ([63] version 4.1.1) as an interactive Python environment with the IPython notebook; numpy and scipy ([64]; version 1.10.4, version 0.17.0) for numerical operations; matplotlib ([65]; version 1.5.1) for data visualization; and pandas ([66]; version 0.17.1) for data analyses were used. The code necessary to reproduce the analyses is available on GitHub (https://github.com/schmelling/reciprocal_BLAST).

Reciprocal BLAST and NCBI

The coding sequences of all entries in the genbank protein database [67], which were labeled as "Complete Genome" or "Chromosome", were downloaded from the NCBI FTP server (version May 2016). These sequences, including the coding sequences of *Synechococcus elongatus* PCC 7942 and *Synechocystis* sp. PCC 6803, were used to construct a custom protein database for the homology search. Further, protein sequences of the 23 clock related proteins (Table S1), from *Synechococcus elongatus* PCC 7942 and *Synechocystis* sp. PCC 6803, respectively, were checked against the entries in the Cyanobase Database to ensure correctness [68] (version May 2016). These 23 protein sequences were used as queries for a search of homologs within the custom protein database, applying the standalone version of BLASTP 2.2.30+ [69] (May 2016) with standard parameter (wordsize: 3, substitution matrix: BLOSUM62). The 10,000 best hits with an e-value of 10^{-5} or lower were filtered for further analyses. The first BLAST run returned circa 65,000 hits for all 23 cyanobacterial proteins combined.
These hits were used as queries for a second reverse BLASTP run, searching for homologs in *Synechococcus* 7942 or *Synechocystis* 6803 genomes using the same parameters as above with an altered e-value of 10. Only hits with the original query protein as best reversal hit were accepted for further analyses, thus minimizing false positive results.
Raw and processed data is available on figshare (https://dx.doi.org/10.6084/m9.figshare.3823902.v2, https://dx.doi.org/10.6084/m9.figshare.3823899.v2).

Testing of co-occurence

Co-occurrence of circadian clock proteins was examined by using the right-sided Fisher's exact test [70]. For each of the 94 cyanobacterial strains, all identified homologous clock genes were gathered into one set. The phylogenetic distribution of cyanobacteria in the NCBI genbank database is very imbalanced. Some genera (e.g. *Prochlorococcus* and *Synechococcus*) are covered better than others. To avoid selection bias, we removed sets with identical combinations of genes, resulting in 69 unique clock systems. Null hypothesis of Fisher's exact test is a pairwise independent distribution of the proteins across all clock systems. P-values were corrected for multiple testing after Benjamini-Hochberg [71] with an excepted false discovery rate of 10^{-2}. We denote that due to the nature of statistical testing, proteins appearing in almost all clock systems are always virtually independent to others. All proteins were clustered according to their corrected p-values.

Microarray analysis

The diurnal expression program of six cyanobacterial strains was probed using microarray time series datasets (Table S3). Unfortunately, the data for the two reported *Synechocystis* 6803 experiments by Labiosa and colleagues [72] and Kucho and coworkers [73] could not be obtained. The study by Toepel and coworkers [74] had to be discarded due to the employed ultradian light cycles (6:6 LD cycles). The *Synechocystis* 6803 datasets, the *Synechococcus* 7942 dataset of Ito and colleagues, and the *Anabaena* sp. PCC 7120 dataset were l2m transformed, while the *Cyanothece* ATCC 51142 datasets are only available after transformation. The two biological replicates of the *Anabaena* sp. PCC 7120 dataset were concatenated for the following analyses [75], similar to the *Synechocystis* 6803 dataset [76]. Expression profiles were smoothed using a Savitzky-Golay lowpass filter, as proposed by Yang and colleagues [77], in order to remove pseudo peaks prior to the detection of periodic genes. Diurnally oscillating expression profiles were detected using harmonic regression analysis [78]. The derived p-values for each gene is based on the assumption of a linear background profile as compared to the sinoidal foreground model. After multiple hypothesis testing correction according to Benjamini-Hochberg [71], all datasets yielded

significantly oscillating genes ($q \leq 0.05$) except for *Microcystis aeruginosa* PCC 7806, with 7 samples the shortest dataset.

Multiple sequence alignment

Multiple sequence alignments were constructed by the standalone version of CLUSTAL Omega 1.2.1-1 [79] using 20 iterations, while only one iteration was used to construct the guide tree (May 2016). The sequences for the alignments were obtained from the processed data generated, as described in the method section "Reciprocal BLAST and NCBI". Afterwards the alignments were adjusted to *Synechococcus* 7942 sequence with Jalview [80] and edited multialignments were used to create WebLogos [81].

Cloning, heterologous expression and purification of Kai proteins

To express either GST-fused or full length KaiC1 proteins from *Synechocystis* sp. PCC 6714, *Nostoc punctiforme* ATCC 29133, *Cyanothece* sp. PCC 7424 as well as KaiC3 from *Cyanothece* sp. PCC 7424, *Microcystis aeruginosa* PCC 7806, *Pycrococcus horikoshii* OT3 PH0833, and *Thermococcus litoralis* DSM5473, the respective *kaiC* genes were amplified by PCR from genomic wildtype DNA. The ORFs and primers are listed in Table S4. Amplified sequences were ligated into BamHI and NotI restriction sites of the plasmid pGEX-6P-1 (GE Healthcare). For *P.horikoshii* and *T.litoralis* KaiC3 the amplified PCR products were ligated into pETDuet1 Vector using BamHI/HindIII and PstI/HindIII restriction enzymes in MCS1 respectively. *Escherichia coli* DH5α or BL21 (DE3) cells were transformed with the resulting plasmids (pGEX-kaiC1-*Sy*6714, pGEX-kaiC1-*Npun*29133, pGEX-kaiC1-*Cy*7424, pGEX-kaiC3-*Cy*7424, pGEX-kaiC3-*Mic*7806, pSVA3151, pSVA3152). For GST-fused expression of KaiA-7942 and KaiC-7942 pGEX derivatives, kindly provided by T. Kondo (Nagoya University, Japan), were used. Expression of GST-Kai proteins occurred at 37 °C and 200 rpm in Terrific broth medium containing 100 μg ampicillin ml^{-1}. GST-KaiA-7942 expression was induced with 1 mM isopropyl β-D-thiogalactopyranoside (IPTG) and carried out overnight, whereas GST-KaiC homologs were expressed for 72 hours without induction. Cells were harvested and lysed in ice-cold extraction buffer (50 mM Tris/HCl (pH 8), 150 mM NaCl, 0.5 mM EDTA, 1 mM DTT, 5 mM $MgCl_2$ and 1 mM ATP). Recombinant GST-Kai proteins were affinity purified using Protino Gluthatione Agarose 4B (Macherey–Nagel) as described in Wiegard and colleagues [55]. During the procedure the GST-tag was removed with PreScission protease in cleavage buffer (50 mM Tris/HCl (pH 8), 150 mM NaCl, 1 mM EDTA, 1 mM DTT, 5 mM $MgCl_2$ and 0.5 mM ATP). Homogeneity of the recombinant proteins was controlled by separating them via SDS-PAGE. If it was not sufficient, proteins were further purified by anion-exchange chromatography using a MonoQ 5/50 GL or ResourceQ column (GE Healthcare). After dialysis in reaction buffer (20 mM Tris/HCl (pH 8), 150 mM NaCl, 0.5 mM EDTA, 5 mM $MgCl_2$, and 1 mM ATP) protein concentration was determined using infrared spectroscopy (Direct detect, Merck Millipore). For full-length KaiC3 homologs, *E. coli* BL21 (DE3) RIL cells were transformed with pSVA3151 (*P.horikoshii*) and pSVA3152 (*T.litoralis*), and grown as preculture overnight at 37 °C in LB medium containing ampicillin (50 μg ml^{-1}) and chloramphenicol (34 μg ml^{-1}). Fresh medium containing antibiotic was inoculated with 0.1 % preculture and grown at 37 °C to an OD_600 of 0.7. After induction with 0.3 mM of IPTG, growth was continued for 16 hours at 16 °C. Cells were collected by centrifugation, frozen in liquid nitrogen, and, after storage at -80 °C, resuspended in 50 ml lysis buffer (50 mM Hepes-NaOH, pH 7.2, 150 mM NaCl) containing Complete EDTA-free protease inhibitor cocktail (Roche) together with DNase I and lysed by sonication. Cell debris were removed by centrifugation at 4 °C for 30 min at 20,000 x g. Further Ni-NTA (Sigma Aldrich, Seelze, Germany) based purification was performed using columns equilibrated in purification buffer (50 mM Hepes-NaOH, pH 7.2, 150 mM NaCl). For the removal of unspecifically bound protein columns were washed with 15 column volumes of equilibration buffer including 10 mM imidazole. *T.litoralis* KaiC3 was eluted in the same buffer with 150 mM imidazole containing equilibration buffer, whereas for *P.horikoshii* KaiC3 the elution was carried out in 20 mM MES pH 6.2, 150 mM NaCl, 150 mM imidazole as this protein is stable in low pH buffer. Further purification of *T.litoralis* KaiC3 was achieved by size exclusion chromatography using Superdex 200 10/300 GL. *P.horikoshii* KaiC3 was incubated at 50 °C for 20 min and centrifuged for 15 min at 10.000 x g. Subsequently, the supernatant was dialyzed overnight against 20 mM MES pH 6.2, 150 mM NaCl buffer. As a quality control, proteins were separated via SDS-PAGE and pure proteins were frozen in liquid nitrogen and kept at -80 °C.

In vitro phosphorylation assays

To investigate KaiA dependent phosphate uptake 12 μg of KaiC-7942, KaiC1-*Sy*6714, KaiC1-*N*29133, KaiC1-*Cy*7424, KaiC3-*Cy*7424, KaiC3-*Mic*7806, KaiC3-*T.lit* or KaiC3-*P.hor* were mixed with 10 μCi γ-P^{32}-ATP in 60 μl Tris reaction buffer (20 mM Tris/HCl (pH 8), 150 mM NaCl, 0.5 mM EDTA, 5

mM $MgCl_2$, 1 mM ATP) in the presence or absence of 6 μg KaiA-7942. 10 μl aliquots were taken after 0, 0.75, 1.5, 3 and 22 hours of incubation at 30 °C and reaction was stopped by adding SDS-sample buffer. Proteins were separated in high-resolution polyacrylamide gels (10 % T, 0.67 % C) by SDS-PAGE (modified from [82]), stained with Coomassie brilliant blue and subjected to autoradiography. Signals were analyzed using a Fujifilm FLA-3000 (FUJIFILM). To analyze in vitro phosphorylation of KaiC3-*T.lit* and KaiC3-*P.hor* at higher temperatures, the recombinant proteins were incubated with 10 μCi γ-P^{32}-ATP in HEPES reaction buffer (50 mM HEPES (pH 7.2), 150 mM NaCl, 5 mM $MgCl_2$, 1 mM ATP) or MES reaction buffer (50 mM MES (pH 6), 150 mM NaCl, 5 mM $MgCl_2$, 1 mM ATP), respectively, at 75 °C. After 0, 5, 10 and 15 minutes 10 μl aliquots were taken and analyzed by SDS-PAGE and autoradiograpy as described above.

Results and Discussion

Most complete set of circadian clock orthologs in Cyanobacteria

Circadian rhythms are reported for many Cyanobacteria and first sequence analyses revealed that the core genes, known from *Synechococcus* 7942, are conserved in almost all cyanobacterial species [4, 53, 55, 83]. Even though daily rhythms seem to be conserved in Cyanobacteria, composition and quantity of corresponding genes on the genome level show high variability [54]. Dvornyk and colleagues first attempted to describe the variety of cyanobacterial core circadian clock systems in 2003 [53]. Since then, the amount and depth of sequencing data increased manifold, which allowed us to perform a detailed analysis of the KaiC-based circadian clock including the input and output pathways. The circadian clock proteins of *Synechococcus* 7942 (Table S1) and the KaiB and KaiC homologs from *Synechocystis* 6803 (Table S1) were the basis of our analysis. Their protein sequences served as base for a reciprocal best hit BLAST analysis. Organisms with at least one homolog to KaiC were retained for further analysis. This stringent filter was essential since KaiC represents the core of the circadian clock in *Synechococcus* 7942. These organisms were grouped by their corresponding genus for a first overview of the homolog distribution (Fig. 1A). We found homologs in Cyanobacteria, Proteobacteria, Archaea, as well as other Bacteria such as *Chloroflexi*. This finding is in good agreement with previous studies [53, 55, 84]. However, our comprehensive study identified a plethora of new bacterial and archaeal genera harboring homologs to the circadian clock genes (Fig. 1A). Nevertheless, Cyanobacteria represent the phylum with the highest degree of sequence similarity and integrity of the system followed by Bacteria, mostly Proteobacteria. In Archaea, homologs to only a fraction of core genes could be identified (Fig. 1A).

Core circadian clock factors KaiB and KaiC beyond the phylum of Cyanobacteria

Four out of seventeen studied factors are exclusively found in Cyanobacteria (Fig. 1A). One of these factors is KaiA, as previously reported in studies with a smaller sample size [53, 84]. However, even some Cyanobacteria, like *Candidatus Atelocyanobacterium thalassa* (previously named Cyanobacterium UCYN-A), and all representatives from the genus *Prochlorococcus* lack *kaiA* (Fig. 1A) [54]. Interestingly, multiple copies of *kaiA* in a single cyanobacterial genome could not be identified. This is of special interest, because we could observe strong sequence length variations for KaiA (Fig. 5A, C) and multiple copies for the other core proteins KaiB and KaiC have been reported (Fig. 1B) [53, 54, 55]. With the *kaiA*-lacking *Candidatus Atelocyanobacterium thalassa* and a *kaiA*-containing cyanobacterial endosymbiont [85] only two out of 94 studied Cyanobacteria do not contain a *kaiB*, whereas a third of the cyanobacterial genera contain multiple copies of kaiB and kaiC, which are also homologs to other known KaiB or KaiC variants (Fig 1B, C). There are a few exceptional cyanobacteria like *Gloeobacter violaceus* [86], which are even lacking the *kaiC* gene and are thus not detected in this analysis due to the previously described filtering criteria. However, all Cyanobacteria having homologs to *kaiB* and *kaiC* are coding for at least one pair of proteins most similar to KaiB1 and KaiC1 from *Synechocystis* 6803 (Fig. 1B, C), which is consistent with previous studies [53, 55]. Only *Cyanothece* sp. PCC 7822 has homologs similar to all three *kaiB* and *kaiC* copies from *Synechocystis* 6803 (Fig. 1C). The majority (67.68%) of the bacterial genera, outside of the cyanobacterial phylum, encode KaiC3-like proteins. Approximately half of them also contain additional KaiC homologs, most similar to KaiC1, KaiC2, or sometimes even both (Fig. 1B). This observation does not hold true for KaiB homologs, here KaiB2 is the major KaiB homolog (75.86%). One group of organisms, all belonging to the phylum of *Bacteroidetes*, stands out in this analysis, because they encode only proteins similar to KaiB2/KaiC2 (Fig. 1B). Archaeal genera show mainly homologs with highest identity to KaiC1 or KaiC3. Furthermore, almost all of the Archaea have multiple copies of the KaiC protein. Whereas only four have homologs to KaiB, which is either similar to KaiB2 or in thermophilic *Methanothermobacter* similar to KaiB1 (Fig. 1B).

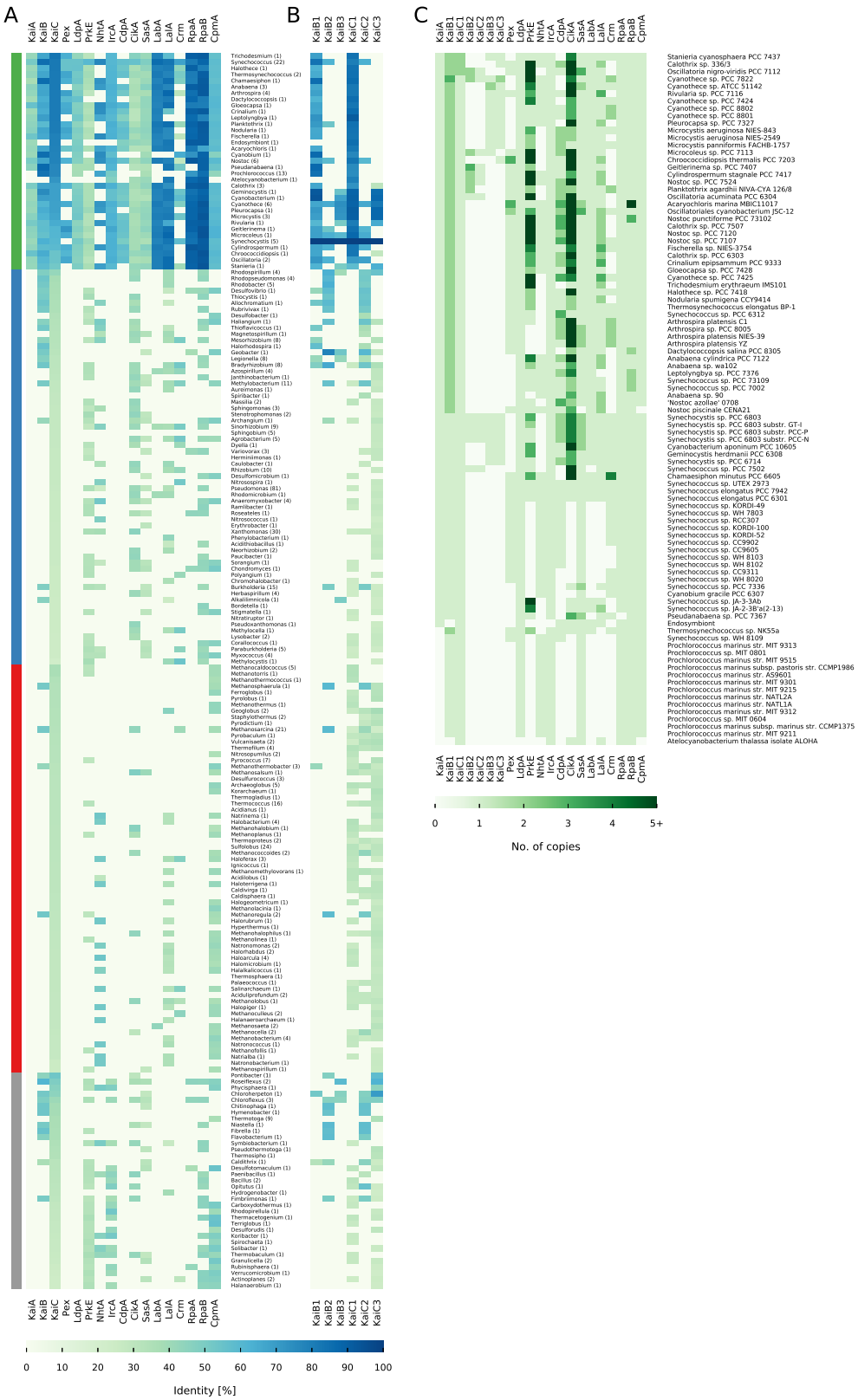

Figure 1: **Distribution of circadian clock proteins**. (**A**,**B**) Shown are the mean sequence similarites of each protein for each genus that contains a KaiC homolog from *Synechococcus* 7942. The genera are sorted by their group and KaiC similarity. The number in parenthesis represents the number of individual genomes per genus. (**A**) shows the mean similarities for the circadian clock proteins of *Synechococcus* 7942. (**B**) shows the mean similarities for the circadian clock proteins of *Synechocystis* 6803. The four taxonomic main groups are highlighted: Cyanobacteria (Green), Proteobacteria (Blue), Archaea (Red), Other (Grey). (**C**) shows the number of homologs for each protein in each cyanobacterium. Copy numbers higher than five were condensed.

Circadian clock factors involved in solely cyanobacterial input pathways

Multiple clock input factors are present in all four taxonomic groups namely PrkE, NhtA, and CikA. The latter also acts in the output of the circadian clock (Table S2, Fig. 1A) [36, 37, 44]. The prevalence of CikA is in good agreement with its tremendous effect on resetting the circadian clock in *Synechococcus* 7942 [37], where its interaction with the Kai complex is mediated by additional proteins [36, 38]. CikA destabilizes in the presence of oxidized quinones and thereby integrates information about the cellular redox state into the oscillator [36]. Naturally, CikA-lacking *Prochlorococcus* are not able to reset their timing mechanism in continuous light [87], pointing at the importance of this factor. However, single mutations in *sasA* can restore circadian properties in *Synechococcus* 7942 *cikA* mutants and it has been suggested that a simpler network with modified interactions of the other clock proteins can exist [88].
The input factors Pex, LdpA, and CdpA are found to be unique for cyanobacteria (Fig. 1A), confirming a previous analysis of *ldpA* [89]. Hence, with Pex, a transcriptional repressor of *kaiA* [34], and LdpA, a redox-sensing protein [20] two factors, which sense the cellular metabolic state of the circadian clock in *Synechococcus* 7942, are missing outside of the cyanobacterial phylum. LdpA is also the only input factor present in the reduced but functional timing system of *Prochlorococcus* [54, 83]. This might indicate the necessity of this factor for the entrainment of the clock [60]. Nevertheless the possibility of a functional clock without Pex and LdpA remains, since LdpA and Pex mutants in *Synechococcus* 7942 are only altered in their period length [90, 35] and Pex is also missing in the KaiA-lacking *Prochlorococcus* and the KaiA-containing *Synechocystis* 6803 (Fig. 1A). In *Synechococcus* 7942 the third unique cyanobacterial input factor, CdpA, influences phase resetting and acts in parallel to CikA [38]. Since CdpA seems to be essential in *Synechococcus* 7942 [38], it likely has a prior role in processes other than phase resetting and is hence dispensable for input pathways in other organisms than Cyanobacteria.
Altogether, the absence of three important input factors outside of the cyanobacterial phylum suggests that other entrainment systems might be used for putative timing systems. In *Rhodobacter sphaeroides*, which displays circadian gene expression rhythms, a histidine kinase is encoded in an operon with *kaiBC* and was suggested as a candidate for transducing the redox signal to KaiBC [91]. Further, the direct entrainment by the ATP/ADP ratio [30, 92] might be the primary mechanism to synchronize the circadian clock with metabolism and the environment.

Central output factors are missing in Archaea and non-cyanobacterial genera

The output pathway of the circadian clock in *Synechococcus* 7942 involves eight proteins (see Conclussion). RpaA serves as a key regulator [43]. Its activity is indirectly modulated depending on the phosphorylation state and the ATPase activity of KaiC [44, 93]. SasA (antagonistically to CikA) connects the core clock to RpaA, which in turn regulates global gene expression, including the *kaiBC* promoter [41, 43, 44, 94]. LabA, CrmA and RpaB are also known to affect RpaA [42, 46, 47, 50, 94]. CpmA modulates *kaiA* expression by an unknown mechanism [49]. In contrast to the unique cyanobacterial input factors, we found none of the eight output proteins exclusively in Cyanobacteria. Homologs of five factors (SasA, CikA, LalA, Crm, CpmA) are present in all four investigated taxonomic groups (Table S2). CpmA is a member of a superfamily essential for purine biosynthesis and thus likely to have orthologs in other organisms [49]. However, RpaA and RpaB are not present in Archaea and SasA is only found in the methanogenic genera *Methanospirillum*, and *Methanosalsum*. Hence, the entire central output pathway is missing in Archaea (Fig. 1A). In addition, orthologs for RpaA are found in only nine non-cyanobacterial genera, questioning whether another transcription factor might read out the putative core timer in other Bacteria (Fig. 1A). A previous BLAST search by Dvornyk and colleagues revealed that SasA homologs in non-cyanobacterial prokaryotes lack the KaiB-like domain [61]. This finding is confirmed in our analysis, indicating that stimulations of SasA homologs by KaiC outside of Cyanobacteria are very unlikely, because interaction occurs via this KaiB-like domain, which adopts a thioredoxin-like fold [39, 95]. Altogether, our analysis reveals that possible circadian clocks of Bacteria and Archaea must use an output pathway that is different from the one described in *Synechococcus* 7942.

Co-occurrence analysis hints at the core module for circadian timing

The previous analysis revealed substantial differences in the composition of the clock components between Cyanobacteria, other Bacteria, and Archaea. Even within Cyanobacteria there is a huge variety in the composition of the potential circadian clocks (Fig. 1C). Cyanobacteria have either a severely reduced timing systems, such as the one in *Prochlorococcus*, a standard system as seen in *Synechococcus* 7942, or an inflated system as found in *Synechocystis* 6803 (Fig. 1C). This trichotomy of systems raises questions about essentiality and pairwise co-occurrence of circadian clock proteins. These questions were answered in a series of right-sided Fisher's exact tests. To avoid systematic biases due to an overrepresentation

of closely related strains [96], we extracted 69 unique combinations of the 21 circadian clock factors as described in materials and methods.

Within these 69 unique systems two factors are always present: (i) KaiC, because we selected for organisms containing at least one KaiC-like protein and (ii) RpaB, which is associated with cell size and circadian gene expression [42, 51]. RpaB competes for promoter binding sites with RpaA and its phosphorylated state is thought to inhibit the phosphorylation of RpaA [47]. Other factors present in the majority ($\geq$ 90 %) of the observed unique clock systems are KaiA, KaiB, LdpA, IrcA, CikA, SasA, RpaA and CpmA. Because of their abundance, most of these factors show no pairwise co-occurence. For example RpaA and RpaB are found in 68 and all 69 clock systems, respectively. Thus their joint presence comes to no surprise. Instead, the finding confirms essential roles in global transcription regulation of Cyanobacteria [97]. With Fisher's exact test we seek to identify gene pairs rather unexpectedly co-occuring in a smaller subset of organisms. Such findings can indicate a common function in the circadian clock system. Only KaiA and CikA, out of the most abundant factors, show significant co-occurence with other factors. In addition, Pex, PrkE, CdpA, LabA, LalA, as well as KaiB2 and KaiC2, and KaiB3 and KaiC3, co-occur significantly (Fig. 2A). Interestingly, all of these factors are missing in *Prochlorococcus*. Additionally, PrkE, CikA, LabA, and LalA are also missing in most marine *Synechococcus* species.

We detected three significant co-occurences between factors of the input pathway (Fig. 2): (i) between CikA and its interaction partner PrkE [38] (ii) between PrkE and CdpA and (iii) between the *kaiA* repressor Pex and CdpA. The first two results are in good agreement with a previous study [38]. Within the output pathway, there is a significant co-occurrence between LabA and its ortholog LalA. Interestingly, we identified several significant co-occurrences between factors of the input and the output pathway. CikA, which functions in the input and output of the clock, co-occurs significantly with LabA, and LalA (Fig. 2). This fits well in the overall picture as CikA and LabA are thought to regulate the activity of RpaA [44, 50]. Additionally, PrkE shows also significant occurrences with both LabA and LalA (Fig. 2). Furthermore, CdpA was found to co-occur significantly with LabA. (Fig. 2) This is of special interest since PrkE and CdpA are only known as interaction partners of CikA, and both are involved in phase resetting, and cell division, respectively [38]. This result, however, hints at potential increased involvement of PrkE and CdpA in the RpaA regulation and supports the view of an integrated network with overlapping interactions of input and output factors [98]. Notably, no co-occurrence of NhtA and LdpA was detected, although it was suggested that NhtA might be involved in assembly of the iron-sulfur cofactor of LdpA [38]. Significant co-occurrences with core factors could only be observed between KaiA and LalA (Fig. 2). However, KaiA shows a strong, but not significant, co-occurrence with CikA ($p = 0.0105$). Lastly, we also found significant co-occurrences between KaiB2 and KaiC2 as well as KaiB3 and KaiC3 (Fig. 2). This indicates two distinct function of the two pairs. In this context it is worth mentioning that *kaiB2*-containing archaeal genomes always encode a KaiC2 homolog.

In summary, we identified a conserved set of factors (Fig. 2), both in input and output that show significant co-occurrences. This set, composed of KaiA, PrkE, CdpA, CikA, LabA, and LalA, is found in Cyanobacteria with a true circadian clock such as *Synechococcus* 7942 and *Synechocystis* 6803 [73] but is missing in cyanobacterial strains with reduced timing mechanisms such as *Prochlorococcus*. This finding hints at the importance of these factors for the functionality of a circadian clock. On the other hand, NhtA and Crm seem to play only a minor or extending role in clock regulation, because they are neither always present nor show significant co-occurrence with other factors.

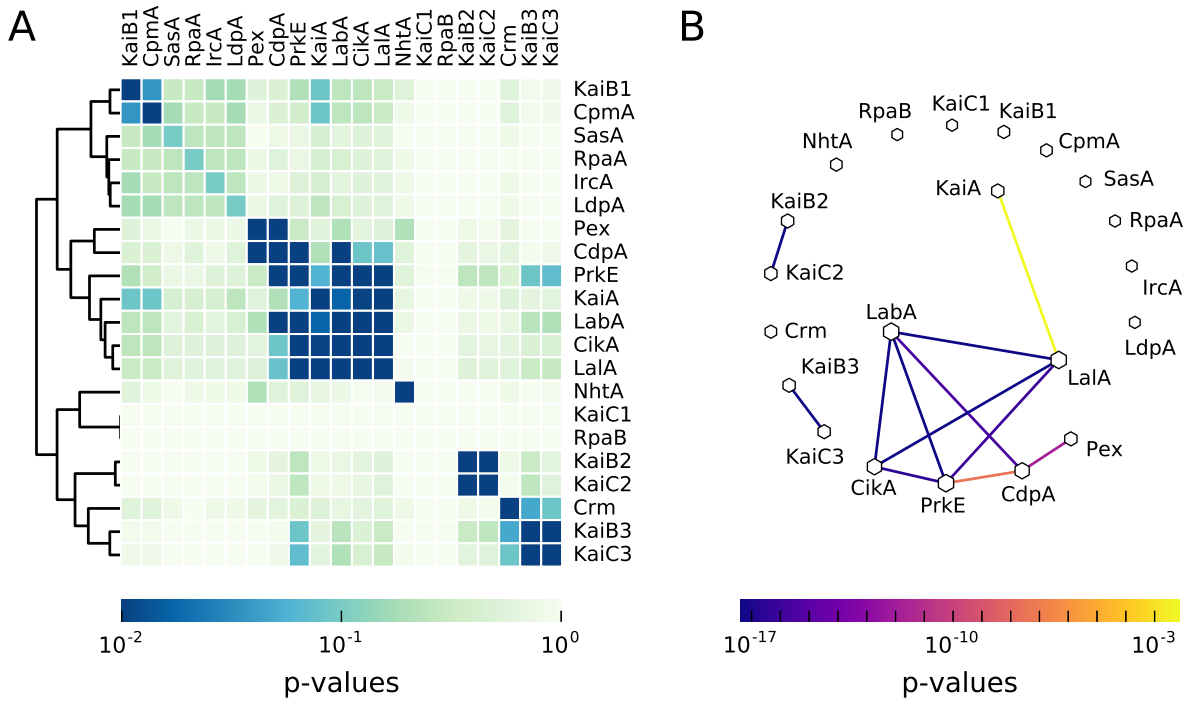

Figure 2: **Co-occurrence of circadian clock proteins in cyanobacteria.** (**A**) The p-values, calculated by pair-wise Fisher's exact tests, are visualized in a heatmap. Only p-values ≤ 0.01 are considered as significant. Proteins are sorted by a hierarchical agglomerative clustering algorithm. (**B**) Network of significant co-occurring circadian clock factors in cyanobacteria, calculated in regard to the results of the pair-wise Fisher's exact test. The line color corresponds to the level of significance. Missing links are those that had a higher p-value than 0.01. Node size is proportional to the degree of that node.

A systematic analysis of circadian expression in Cyanobacteria

Genome-wide time-resolved expression measurements in a range of cyanobacterial strains have repeatedly indicated substantial fractions of genes with circadian regulation patterns [99, 100, 101]. Considering that all Cyanobacteria share the challenge of a photoautotrophic lifestyle, which requires major changes in the metabolism between day and night, one might expect a common transcriptional regulatory pattern. Thus, we compared a total of nine published microarray time-series datasets of different cyanobacterial strains under constant light or diurnal light conditions (for details see Table S3, Supplement), which were available and applicable for this analysis. Not all of the chosen microarray experiments were conducted under constant light conditions, which leads to a combination of circadian-clock regulated and light-induced genes. We therefore refer to genes with oscillating expression as diurnally regulated instead of circadian. For allowing a direct comparison, we reprocessed the raw-data and subjected the resulting expression time series to a harmonic regression oscillation detection. This method assumes a sinusoidal shape of circadian expression profiles and uses linear expression profiles as background, yielding estimates of the peak phase and amplitude of each gene.

In a first step we compare biological replicate datasets to establish the reproducibility of strain-specific circadian expression programs. Similarity between two circadian expression programs was established using the circular correlation coefficient ρccc as described by Jammalamadaka and Sarma [102] applied to estimated peak expression phases. The following analyses were limited to genes with oscillating expression profiles in both compared datasets since only in these cases phase and amplitude estimates are meaningful descriptors. Direct comparison of the oscillation phases and amplitudes indicates good reproducibility between two respective measurements in form of statistically significant elevated correlation of the circadian expression patterns in *Synechococcus* 7942 (ρccc = 0.61, p $\ll$ 0.01), *Synechocystis* 6803 (ρccc = 0.31, p $\ll$ 0.01), and *Cyanothece* sp. ATCC 51142 (ρccc = -0.51, p $\ll$ 0.01) (Figure 4 top row). While the *Synechocystis* 6803 datasets show significant similarity, the correlation is diminished by the distinct concentration of expression phases during the day in the beck14 dataset compared to the leh13 measurements. Interestingly, both *Cyanothece* sp. ATCC 51142 datasets exhibit a good agreement of peak expression phases with large early day and early night clusters, but the large negative correlation

emphasizes the presence of a significant number of anti-phasic gene pairs. The corresponding oscillation amplitude values exhibit high statistically significant correlations (ρccc > 0.77) for all three datasets (Fig. 3 bottom row).

This observation motivated the second step of the analysis, the comparison of expression patterns between different cyanobacterial strains. To facilitate this comparison, the prediction of homologous genes in the cyanobacterial clade by Beck and colleagues [103] was used as starting point. We focused first on the set of genes with oscillating expression patterns in all datasets. The estimate of the core diurnal genome based on the presented data collection spans 95 genes (Table S5), which are mostly involved in central metabolic processes. Most strikingly, 18 out of 64 genes annotated by Cyanobase [68] as "ribosomal protein" genes in *Synechocystis* 6803 fall into the core diurnal set, furthermore seven out of 27 genes annotated with "photosystem II". The remaining diurnally expressed genes are found interspersed across the metabolic network. While genes coding for parts of photosystem, the RNA polymerase, or the ribosomal proteins can be expected to serve important roles in the adaptation to photic and aphotic phases, this analysis ascribes similar importance to other metabolic processes e.g. in the repair of UV-damaged Photosystem II centers (*slr1390*) [104], in the phosphate transport system (*pstB2*), pyrimidine and arginine biosynthesis (*sll1498*), or the glycolysis/gluconeogenesis via the fructose-bisphosphate aldolase (*sll0018*). By excluding genes, which possess light-inducible expression, marker gene candidates for a working clock can be derived from the core diurnal genome, such as the light-independent protochlorophyllide reductase subunit ChlB and the fructose-bisphosphate aldolase *fbp*. Indeed, *fbp* also shows circadian expression patterns in *Clamydomonas reinhardtii* [105] and *Arabidopsis thaliana* where its late-night peaks may reflect the great importance of these aldolases in higher plants for the mobilization of plastidic starch [106]. Particularly in higher plants, the mobilization of starch, the conversion into sucrose, and its transport to other parts of the plant occur mainly at night.

The group of 15 "hypothetical protein" genes in the core diurnal genome constitutes an excellent candidate set for novel clock-driven genes in strains with a working core clock. Interestingly, several of these genes are implicated with cell division, such as the YlmG-related hypothetical gene *ssl0353*, which is required for proper distribution of nucleoids in cyanobacteria and chloroplasts [107]. Similarly, the hypothetical protein *slr1577* is suggested to function in the separation of chromosomes during cell division (Uniprot entry P74610). For the gene *slr1847* (Uniprot entry P73057) a DNA binding capability is suggested, which could therefore regulate expression, aid nucleoid organization, or protect the DNA.

The core clock genes *kaiA*, *kaiB*, and *kaiC* are a good starting point for a detailed comparative expression analysis. Only the *kaiB1* is significantly oscillating in all considered datasets. Interestingly, the *kaiA* gene features only very low amplitude expression oscillations and is arrhythmic in the vijayan09 *Synechococcus* 7942 dataset. The expression phases vary from dawn (*Microcystis aeruginosa* PCC 7806) to morning (*Synechocystis* 6803), over midday (stoeckel08 dataset of *Cyanothece* sp. ATCC 51142), and dusk (ito09 dataset of *Synechococcus* 7942), into night (*Anabaena* sp. PCC 7120). The observed expression phases of *kaiB1* are comparable to those of *kaiA*, but with significantly larger amplitude in *Synechococcus* 7942 datasets. The phase of the *kaiB1* homolog in *Prochlorococcus marinus* MED4 peaks before dawn, comparable to *Anabaena* sp. PCC 7120. The *kaiC1* expression phases and amplitudes match those of *kaiB1*, with the notable exception of *Cyanothece* sp. ATCC 51142 for which antiphasic late-night peaks are observed. In *Prochlorococcus marinus* MED4, *kaiC1* peaks during the early night in contrast to the late night phase of *kaiB1*.

Many aspects agree well with previous knowledge. In *Synechococcus* 7942, the core clock genes *kaiB* and *kaiC* are arranged in the *kaiBC* operon resulting in similar expression patterns [108, 109, 110], while *Cyanothece* sp. ATCC 51142 features the *kaiAB1C1* operon [111]. Interestingly, *Cyanothece* sp. ATCC 51142 features consistent anti-phasic expression of *kaiB1* and *kaiC1* whereas the remaining strains show co-expression, hinting at *Cyanothece*-specific post-transcriptional regulation of *kaiB1* or *kaiC1*. Oscillations in the *kaiA* gene expression, as reported by Ishiura and colleagues, feature small expression amplitudes compared to *kaiB* and *kaiC* [112]. In fact, *kaiA* consistently falls below the threshold of 2-fold expression change for the classification as circadian oscillator, which is commonly employed in microarray studies.

In the second step we generalized the detailed analysis of expression phases as presented for the core diurnal genome. We applied the circular correlation measure to all possible combinations of expression datasets. The resulting distribution reveals a clear separation between pairs of biological replicate datasets, featuring large numbers of shared oscillating genes and more extreme correlations (Fig. 4 red), and pairs of different cyanobacterial strains with fewer shared oscillating genes and much less extreme correlation coefficients (Fig. 4 blue). The only exception to this separation is the *Cyanothece* sp. ATCC 51142 dataset stoeckel08 (Fig. 4 green), which shares many oscillating genes with both *Synechocystis* 6803 datasets (leh13, beck14). The corresponding correlation coefficients are, however, similarly small compared to other inter-strain pairs. The full set of pairwise phase comparisons, which underlay this

analysis are shown in Figure S1. This result indicates that the diurnal peak expression phase is not preserved amongst homologous genes in the cyanobacterial clade but might instead be tuned according to the metabolic gene outfit and the environmental needs of the respective strain.

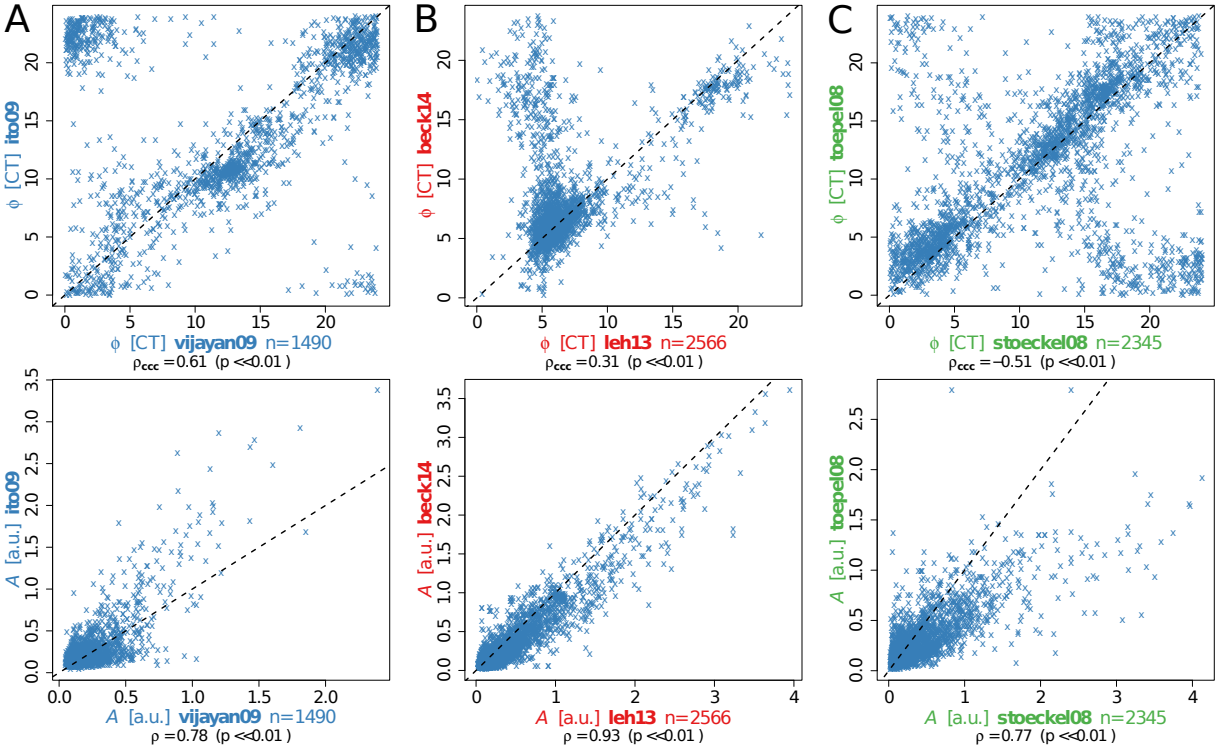

Figure 3: **Phase and amplitude reproducibility of diurnal genes within cyanobacterial strains.** Comparison of expression phase φ (top, [CT]) and amplitude A (bottom, [a.u.]) of diurnal genes shared between independent datasets of the same cyanobacterial strain. (**A**) *Synechococcus* datasets of Vijayan and colleagues [113] (x-axis) and Ito and coworkers [101] (y-axis), (**B**) *Synechocystis* in the datasets of Lehmann and coworkers [76] (x-axis) and Beck and colleagues [103] (y-axis), and (**C**) *Cyanothece* in Stöckel and colleagues [114] (x-axis) and Toepel and colleagues [115] (y-axis). The number of genes n found to oscillate significantly and the corresponding Pearson correlation coefficient ρ between φ and A is provided below each panel, followed by the respective p-value for ρ differing from 0. For φ, the circular correlation coefficient ρccc is also provided. Axis labels are shown in the strain-specific color.

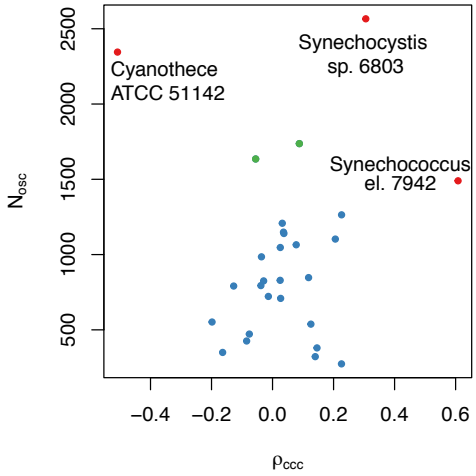

Figure 4: **Expression phase similarity versus number of shared oscillating genes.** The comparison is carried out between all dataset combinations. Biological replicate experiments available for three cyanobacterial strains show a large number of shared oscillating genes as well as large correlation coefficients (red). The comparison of the stoeckel08 dataset (*Cyanothece*) with both *Synechocystis* datasets is marked in green. The remaining comparisons between various strains are shown in blue.

Core circadian clock proteins, KaiA, KaiB, KaiC, vary in number and length

The previously observed diversity of circadian clocks within Cyanobacteria, and between other Bacteria and Archaea prompted further sequence analyses of the core clock proteins KaiA, KaiB and KaiC. Length comparisons gave rise to some new features of variations between the core factors of the circadian clock (Fig. 5). As described in the preceding, our BLAST analysis detected KaiA exclusively in Cyanobacteria. Interestingly, we could distinguish three subtypes of KaiA. While the sequence length of most KaiA is around 300 amino acids (AA) (*Synechococcus* 7942: 284 AA) some stains have shortend homologs with a length of roughly 200 and 100 AA, respectively (Fig. 5A, C). Truncated KaiA proteins are almost exclusively found in members of the order *Nostocales*. Similar results were reported previously by Dvornyk and colleagues, who also observed a higher degree of polymorphism for the *kaiA* gene in comparison to *kaiB* and *kaiC* [52, 53]. Multiple alignments of the KaiA proteins (Fig. S2) verified that the truncated KaiA proteins have a shortened N-terminal sequence, which functions in the complete protein as the amplitude amplifier [116]. However, all of these KaiA orthologs contain the C-terminal part important for clock oscillation [116].

Distribution of the KaiB protein length reveals two distinct groups. The KaiB homologs are either as long as the one from *Synechococcus* 7942 (102 AA) or about 250 AA in length (Fig. 5B, C). *Microcoleus* sp. PCC 7113 even has a KaiB with the length of 381 AA. KaiB homologs with query length are present in all four groups (Fig. 5B). Elongated KaiB proteins are mainly present in Cyanobacteria in the subclass *Oscillatoriophycideae* and the order *Nostocales*, specifying findings of Dvornyk [53] (Fig. 5C). BLAST analyses using the KaiB homologs from *Synechocystis* 6803 revealed that elongated variants are most similar to KaiB1. Elongation via concatenation of two KaiB was ruled out by visually inspecting alignments with two artificially concatenated KaiB1. Instead the elongated KaiB1 have a ∼150 AA N-terminal extension. BLAST searches of the N-terminal region showed no homologous sequences in other organisms than Cyanobacteria, and no putative conserved domains could be identified. However, the N-terminal part is highly conserved within the Cyanobacteria having this KaiB variant. Interestingly, those *Nostocales* with an elongated KaiB1, also show a truncated KaiA.

The KaiC protein from *Synechococcus* 7942 is 519 AA in length and is build up by two domains, the CI and CII domain, which have a high similarity and are connected by a linker-domain [117, 118]. The C-terminal CII domain of KaiC comprises the interaction sites with KaiA as well as the specific phosphorylation sites [29, 119, 120, 121]. KaiC homologs were detected with lengths varying between 101 AA and 741 AA (Fig. 5A, B). There are KaiC homologs in Archaea representing the whole observed length spectrum of KaiC, whereas bacterial KaiCs are almost always about 500 AA in length (Fig. 5A, B). Furthermore, in Bacteria and Archaea KaiB (and KaiA in Cyanobacteria) is only found when a 'full-length' KaiC is present. In these bacterial organisms the length of KaiC is almost constant, regardless of the length of KaiA and KaiB homologs (Fig. 5A, B). However, in KaiB-possessing Archaea additional shorter KaiC homologs are found (Fig. 5B).

Moreover, the length distribution of KaiC revealed a substantial amount of KaiC homologs with a length of circa 250 AA, which is approximately the length of one KaiC domain. This KaiC variant is mainly found in Archaea, but also in a few Bacteria. In these bacterial species no KaiB homolog could be identified. Shorter KaiC homologs do not contain the important phosphorylation sites for maintaining the oscillator function. Therefore, they might not restore the full functionality of the *Synechococcus* 7942 KaiC, but can rather answer questions about the evolution of KaiC [53]. Regarding the evolution of KaiC, two valid hypotheses exist, both of which state that KaiC arose from a shorter ancestral *recA* gene followed by a gene duplication and fusion. However, on the one hand Leipe and colleagues [122] hypothesize that an ancestral single-domain KaiC originated in Bacteria, was transferred into Archaea, where its two-domain version originated, and a second lateral transfer event introduced the double domain KaiC into cyanobacteria. On the other hand Dvornyk and coworkers [53] argue in a follow up study that KaiC has to be of cyanobacterial origin. Given the amount of new genomic data further studies would help to unravel the evolutionary history of KaiC.

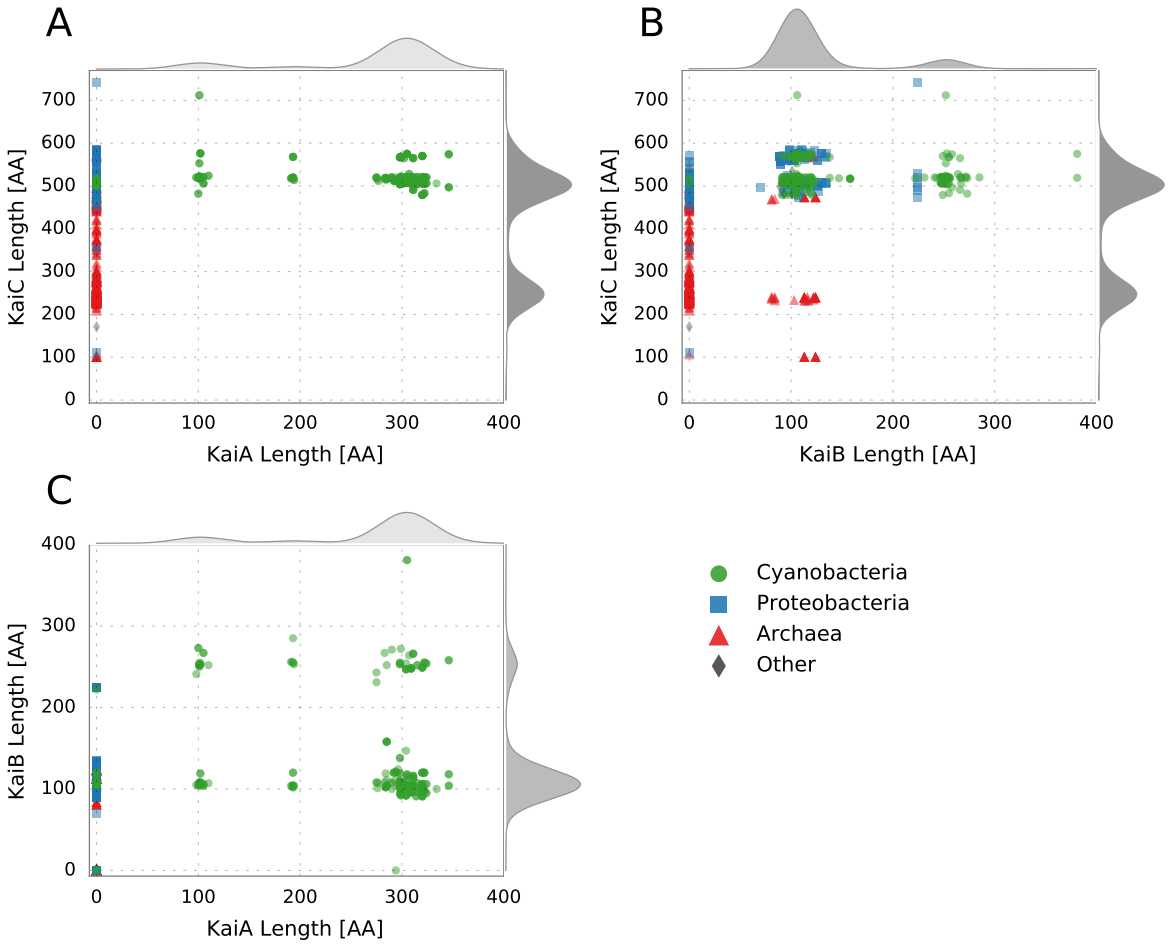

Figure 5: **Protein length distribution of the circadian clock factors from *Synechococcus* 7942 and its homologs.** The four taxonomic main groups are highlighted: Cyanobacteria (Green), Proteobacteria (Blue), Archaea (Red), Other (Grey). The curves outside of the plot represent the cumulative density distribution of the respective protein. (**A**) KaiC length distribution in dependency of the KaiA length. (**B**) KaiC length distribution in dependency of the KaiB length. (**C**) KaiB length distribution in dependency of the KaiA length.

Conserved motifs and activities in the cyanobacterial KaiC subgroups

For KaiC2 homologs outside of the cyanobacterial phylum an involvement in stress response (*Legionella pneumophila* [123]) and adaptive growth under rhythmic conditions (*Rhodopseudomonas palustris* [57]) has been demonstrated. Both proteins display autophosphorylation and KaiC2 from *Rhodopseudomonas palustris* shows elevated ATPase activity [57]. Nevertheless, the function of cyanobacterial KaiC2 and

KaiC3 homologs remains unclear. We already demonstrated that KaiC2 and KaiC3 from *Synechocystis* 6803 displays kinase activity, which is independent of KaiA, whereas KaiC1 behaved like its *Synechococcus* 7942 ortholog [55]. Those activities could also be predicted from the C-terminal amino acid sequences [55]. To test whether general features of the three KaiC subgroups can be predicted, multiple alignments of the cyanobacterial KaiC1, KaiC2 and KaiC3 sequences were constructed. A WebLogo analysis revealed that relevant motifs for phosphorylation and dephosphorylation in the CII domain are highly conserved. The ATP-binding Walker Motif A (P-loop in Fig. 6A, GXXXXGKT, [112, 124, 125]) is present in all three KaiC subgroups. Strikingly, the respective sequence of KaiC-7942 (GATGTGKT) shows almost no modifications in KaiC1 and KaiC3 proteins. Furthermore, catalytic glutamates (EE in Fig. 6A, [126, 127]), the R-finger contacting the γ-phosphate of ATP [128], and the truncated Walker motif B (WalkerB in Fig. 6A, [26, 112, 125]), were found in all cyanobacterial KaiC subgroups. Notably, the arginine residue of the *Synechococcus* 7942 Walker B motif is not conserved in KaiC2 homologs. Serine and subsequent threonine are the dominant phosphorylation sites in KaiC1 and KaiC3 proteins, like S431 and T432 in KaiC-7942 [119, 120], whereas KaiC2 homologs display two serine residues. In some KaiC3 homologs a tyrosin is present as second phosphorylation site. T426, which is important for dephosphorylation of KaiC-7942 [119, 126, 129, 130], is also highly conserved. Therefore, phosphorylation and dephosphorylation via autokinase [26], ATP synthase and ATPase activity [27, 126], respectively, are very likely for all cyanobacterial KaiC homologs. The same holds true for the N-terminal ATPase activity: We observed high conservation of the Walker motif A (P-loop in Fig. 6A), the catalytic glutamate residues (EE in Fig. 6A) and the R-finger in the CI domains of all cyanobacterial KaiC subgroups. The presence of the R-linker in CI domains of KaiC1 and KaiC3 homologs indicate a structural coupling of the N-terminal CI and the C-terminal CII-domain as it was demonstrated for *Thermosynechococcus* KaiC [128], whereas KaiC2 homologs lack the R-linker in CI.

KaiC homologs from the genus *Prochlorococcus* were classified as KaiC1 orthologs in our BLAST analysis (Fig. 1B). However, *Prochlorococcus* strains do not contain KaiA, and KaiC from *Prochlorococcus marinus* MED4 was demonstrated to phosphorylate independently of KaiA due to a modified A-loop sequence [83]. Therefore, we compared WebLogos of the A-loop sequence [28] for KaiC1 orthologs from Cyanobacteria with and without KaiA (Fig. 6A). The most obvious difference to KaiC1 proteins from cyanobacterial strains with KaiA is the presence of neutral glutamine in the second position, instead of a positively charged arginine. In KaiC2 and KaiC3 Weblogos motifs are even less conserved as already described for the *Synechocystis* 6803 representatives [55]. Interestingly, the KaiC3 WebLogo motif does not display any charged residue anymore. The absence of A-loop residues that are important to keep it in the buried state [28] indicate that the KaiA-independent phosphorylation is characteristic for KaiC2 and KaiC3 homologs. This is supported by the low conservation of the 438-444-loop and/or the 422-loop, which are part of the interaction network that mediates inhibition of phosphorylation by the buried A-loops in *Synechococcus* 7942 [28, 118]. The high conservation of these loops in KaiA-lacking strains remains enigmatic.

To test this hypothesis, phosphate uptake as an exemplary KaiC activity was analyzed for representative cyanobacterial KaiC1 and KaiC3 proteins by incubation with γ-P^{32}-ATP at 30 °C in the presence and absence of *Synechococcus* 7942 KaiA (KaiA-7942) The well-studied KaiC from *Synechococcus* 7942 (KaiC-7942) served as control. As demonstrated in Figure 6B and 6C all recombinant KaiC proteins incorporated phosphate over time. The intrinsic kinase activity of KaiC1 homologs from *Nostoc punctiforme* ATCC 29413 (KaiC1-N294133), *Synechocystis* sp. PCC 6714 (KaiC1-Sy6714), and *Cyanothece* sp. PCC 7424 (KaiC1-Cy7424) was stimulated by KaiA-7942, similar to KaiC-7942 (Fig. 6B). As expected, KaiA had no effect on autophosphorylation of KaiC3 from *Cyanothece* sp. PCC 7424 (KaiC3-Cy7424) and *Microcystis aeruginosa* PCC 7806 (KaiC3-Mic7806, Fig. 6C). To extend the analysis to non-cyanobacterial proteins, KaiC3 from the hyperthermophilic Archaea *Thermococcus litoralis* (KaiC3-T.lit) and *Pyrococcus horikoshii* (KaiC3-P.hor), which show optimal growth at 85 °C and 98 °C [131, 132], were analyzed in a similar way. Again both recombinant KaiC3 proteins displayed phosphorylation at 30 °C, which was independent of KaiA-7942 (Fig. 6C). Incubation at 75 °C indicated that the two archaeal KaiC3 homologs display kinase activity also at high growth temperatures (Fig. 6D). Hence kinase activity of KaiC proteins seems to be well-conserved, independent of the growth conditions of the strains, they are originating from.

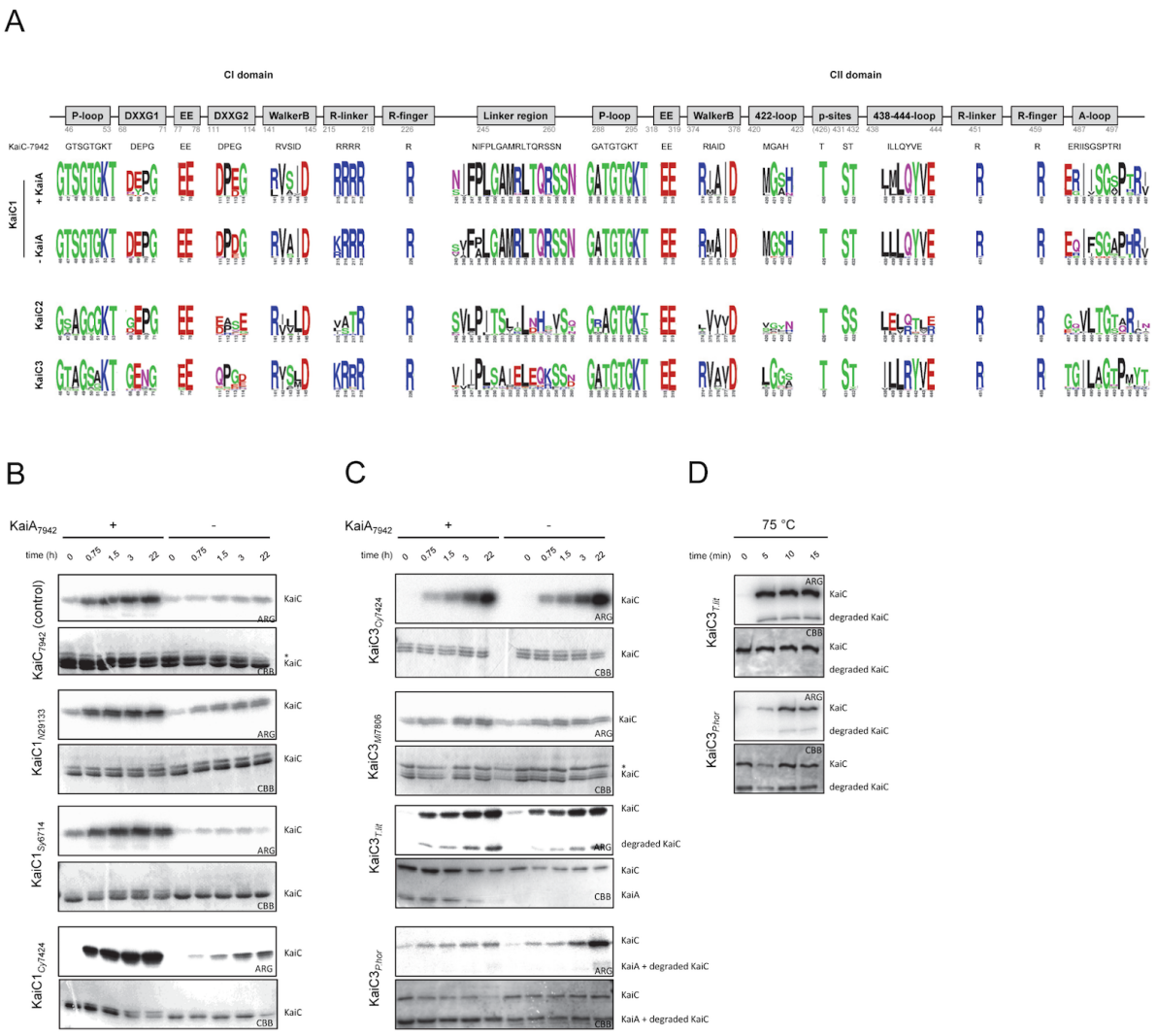

Figure 6: **Activity of diverged KaiC homologs.** (**A**) Conservation of important motifs in cyanobacterial KaiC1, KaiC2 and KaiC3 homologs based on a WebLogo analyses. Motifs in KaiC1 homologs are displayed for proteins from organisms, which encode a KaiA protein or lack KaiA, respectively. Numbers indicate the residues in KaiC-7942. Properties of the residues are displayed as follows: polar (green), neutral (purple), basic (blue), acidic (red), and hydrophobic (black). (**B**,**C**) Phosphate uptake analyses of selected KaiC1 (**B**) and KaiC3 (**C**) homologs at 30 °C in dependence of KaiA. KaiC proteins were incubated with or without KaiA-7942 in the presence of γ-P^{32}-ATP. After 0, 0.75, 1.5, 3, and 22 hours samples were separated via SDS-PAGE, stained with Coomassie (CBB), and subjected to autoradiography (ARG). The asterix indicates a contaminating protein. (**D**) Kinase activity of archaeal KaiC3 homologs at 75 °C. KaiC3 proteins from *Thermococcus litoralis* and *Pyrococcus horikoshii* were incubated with γ-P^{32}-ATP and autophosphorylation was analyzed after 0, 5, 10, and 15 minutes. Shown are the Coomassie stained proteins (CBB) and autoradiography (ARG).

Conclusion

A core module for circadian regulation

Our analysis of 11,264 genomes clearly demonstrates that components of the *Synechococcus* 7942 circadian clock are present in various bacteria and archaea. However, the frequency of Kai-clock related proteins is highest in Cyanobacteria. In fact KaiA, Pex, LdpA, and CdpA are exclusive to organisms of this phylum. In other organisms, e.g *Rhodobacter sphaerodies*, reduced KaiBC-based clock systems are likely able to drive circadian oscillations [91]. An even simpler system solely dependent on KaiC might enable diurnal rhythms in *Haloferax volcanii* [59], probably using the ATP/ADP ratio for clock entrainment. Predictions for KaiC activities based on sequence alignments and motif analyses were validated through a series of biochemical experiments. We confirmed ATPase and kinase activity for 'full-length' KaiC proteins composed of one CI and one CII domains, even in organism without *kaiA* or *kaiB*. KaiA from *Synechococcus* 7942 enhanced KaiC-phosphorylation only in strains naturally possessing a *kaiA* gene.

Our co-occurrence analysis hints to a conserved extension set for circadian regulation, which is present in cyanobacteria with observed circadian behavior and absent in cyanobacteria having a diurnal, hourglass-like lifestyle only (see also Fig. 7). A diurnal core set, which is important to enable an hourglass-like timing system that resets every day, might be composed of KaiB, KaiC, LdpA, IrcA, SasA, RpaA, RpaB, and CpmA. However, our identified circadian core set, which potentially enables a selfsustained clock, consists of KaiA, the two input factors CdpA, and PrkE as well as the input and output factor CikA, and the output factors LabA, and LalA.

The systematic comparison of microarray timeseries datasets indicates that the diurnal peak expression phase is not conserved amongst homologous genes in the cyanobacterial clade. Instead, the expression phase may be tuned according to the gene outfit and varying environmental needs. The analysis yielded a set of 95 genes in the core diurnal genome, which can be considered critical for the adaptation to day and night. Particularly the subset of non-light induced genes are prime candidates for circadian clock marker genes. This set furthermore contains several hypothetical genes, which are interesting candidates for novel clock-driven genes.

The gained insights about the diversity within the composition of the components involved in the circadian protein clock as well as the diversity on the sequence level of the core factors call for further modification and simplification of the clock. The exponential increase of molecular tools for synthetic applications in recent years sets the stage for such ambitious projects. A future goal could be the reduction of complexity by removing as much factors as possible so that an integration of a circadian clock in synthetic and industrially valuable organisms becomes feasible.

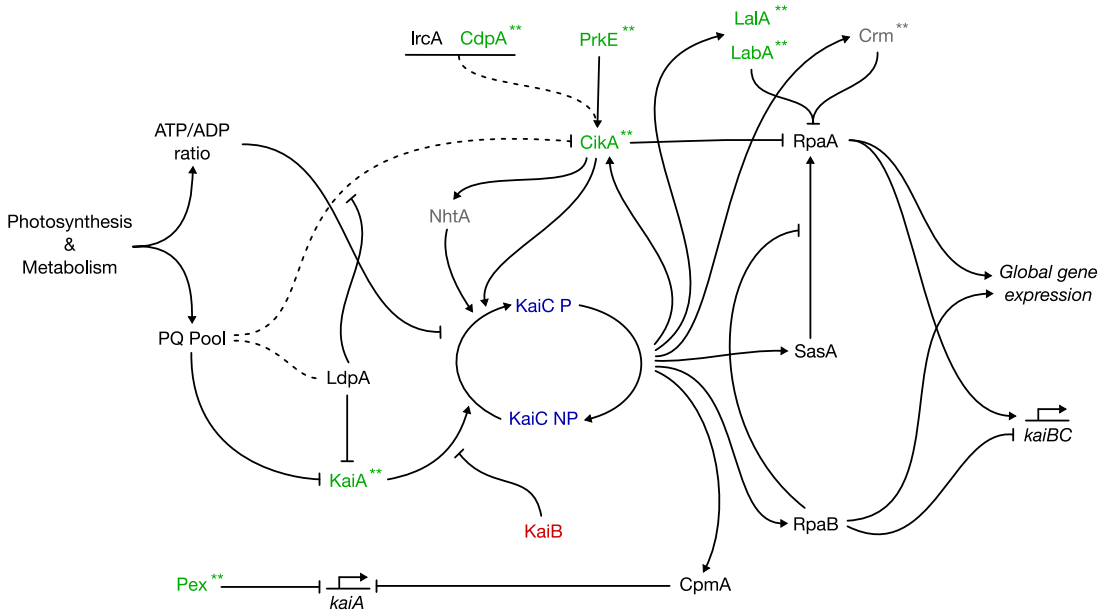

Figure 7: **Schematic overview of the circadian clock in *Synechococcus* 7942.** Shown are the protein interactions based on [20, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51]. The plastoquinone pool and ATP/ADP ratio serve as input signals to entrain the circadian clock with the environment and metabolic state of the cell. The signals are recognized by the input factors LdpA and CikA and the core clock factors KaiA and KaiC. Other input factors are Pex, and the CikA interaction partner NhtA, PrkE, IrcA, and CdpA. The stimulating effect of KaiA on KaiC is antagonized by KaiB. The output of the clock is comprised by SasA, CikA, RpaA, RpaB, LabA, LalA, Crm, and CpmA. SasA interacts with KaiC and further phosphorylates RpaA, whereas CikA acts as a phosphatase on RpaA. RpaA is also regulated by LabA, Crm, and RpaB. RpaA together with RpaB controls global gene expression as well as the expression of the *kaiBC* cluster. CpmA is a transcriptional regulator of *kaiA*. The seven most interconnected factors we found in our co-occurrence analysis are highlighted in green. Factors colored in light grey showed no co-occurrence in the Fisher's exact test and are present in $< 90\%$ of all observed systems. Asterisks (**) indicate the factors missing in *Prochlorococcus*.

Supplemental Material

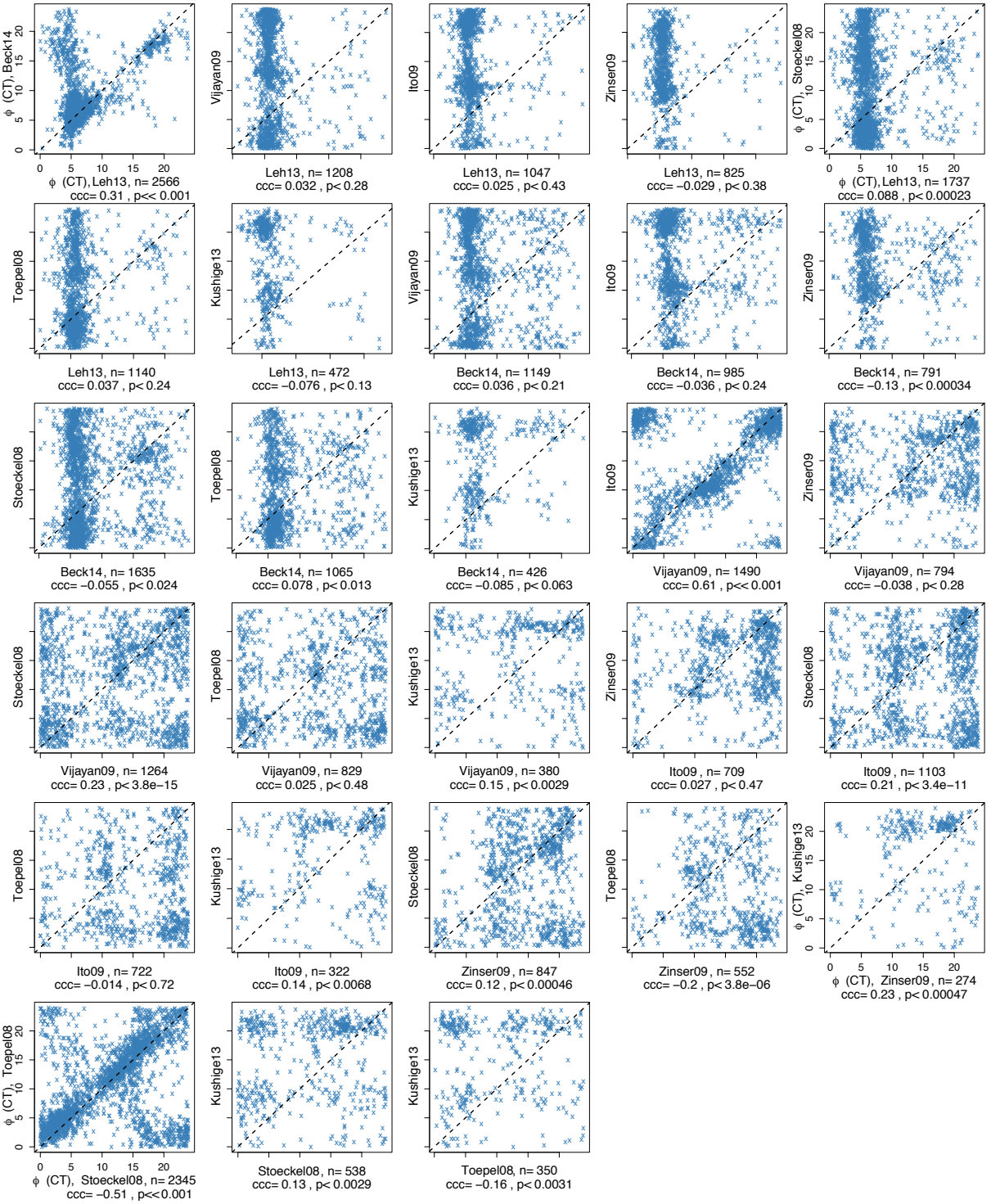

Figure S1: **Expression phase similarity across cyanobacterial strains by pairwise comparison between available datasets.** All possible pairwise combinations of available circadian datasets are compared with respect to peak expression phase, considering only genes, which oscillate significantly (fdr < 0.05) in both datasets. While phases are compared directly for same-strain combinations, gene pairs across different strains are derived via homology prediction. The respective dataset is shown on each axis, the count of homologous genes significantly oscillating is provided with the x-axis label, together with the circular correlation coefficient ρccc and the resulting p-value.

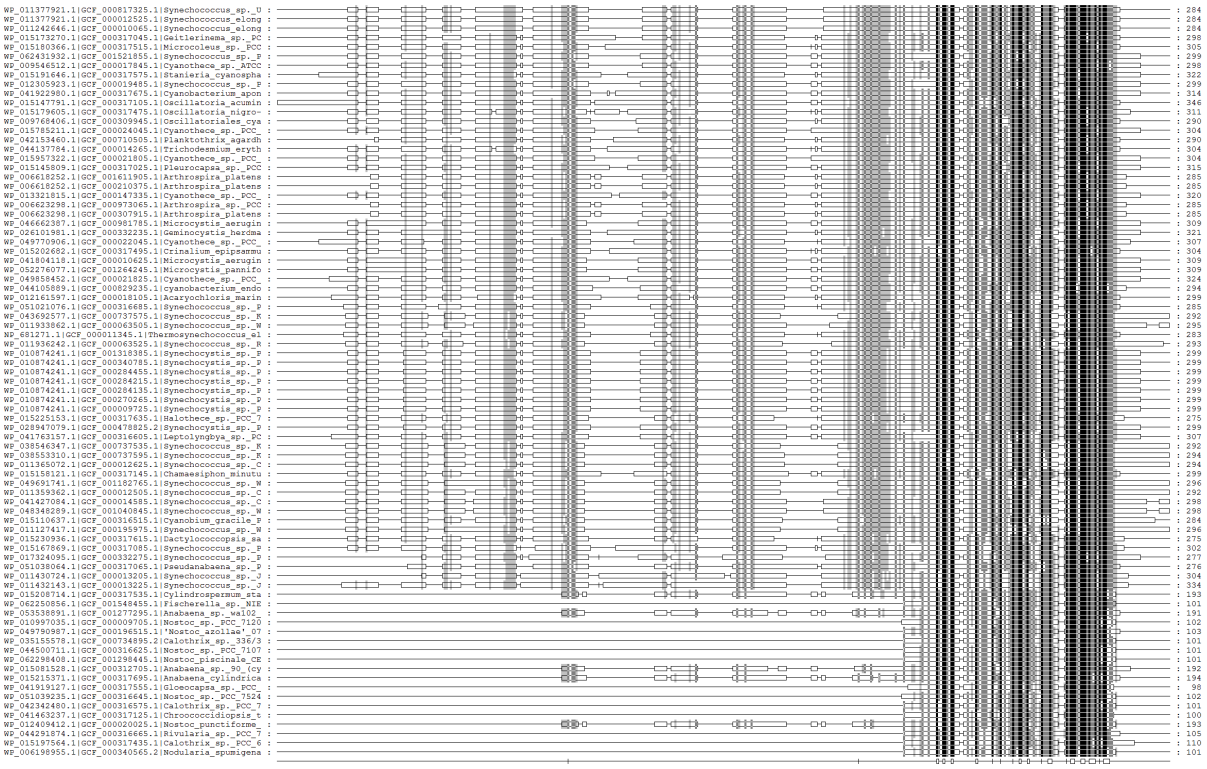

Figure S2: **Multiple sequence alignment of KaiA.** Conserved regions are highlight: ≤100% (black), ≤80% (grey), ≤60% (light grey), ≤40% (white).

Table S1: **Proteins of the circadian clock used as queries for the reciprocal BLAST analysis.**

Organism Name	Function	Protein Name	Gene ID
Synechococcus elongatus PCC 7942	Core	KaiA	Synpcc7942_1218
		KaiB	Synpcc7942_1217
		KaiC	Synpcc7942_1216
	Input	Pex	Synpcc7942_0677
		LdpA	Synpcc7942_0624
		NhtA	Synpcc7942_2160
		PrkE	Synpcc7942_0600
		IrcA	Synpcc7942_2383
		CdpA	Synpcc7942_1604
	Input/Output	CikA	Synpcc7942_0644
	Output	SasA	Synpcc7942_2114
		LabA	Synpcc7942_1891
		LalA	Synpcc7942_1143
		Crm	Synpcc7942_0096
		RpaA	Synpcc7942_0095
		RpaB	Synpcc7942_1453
		CpmA	Synpcc7942_1168
Synechocystis sp. PCC 6803	Core	KaiB1	slr0757
		KaiC1	slr0758
		KaiB2	sll1596
		KaiC2	sll1595
		KaiB3	sll0486
		KaiC3	slr1942

Table S2: **Distribution of *Synechococcus* 7942 based circadian clock proteins in the four main groups.** The percentage indicates the coverage of the orthologs per group.

Group	Core Clock	Input Pathway	Output Pathway
Cyanobacteria	KaiA (94.29%), KaiB (94.29%), KaiC (100%)	Pex (62.86%), LdpA (97.14%), CikA (94.29%), NhtA (48.57%), PrkE (85.71%), IrcA (97.14%), CdpA (74.29%)	CikA (94.29%), SasA (100%), LabA (88.57%), LalA (91.43%), Crm (45.71%), RpaA (97.14%), RpaB (100%), CpmA (97.14%)
Proteobacteria	KaiB (28.13%), KaiC (100%)	CikA (56.25%), NhtA (26.56%), PrkE (51.56%)	CikA (56.25%), SasA (32.81%), LabA (4.69%), LalA (39.06%), Crm (20.31%), RpaA (7.81%), RpaB (48.44%), CpmA (21.86%)
Archaea	KaiB (6.06%), KaiC (100%)	CikA (12.12%), NhtA (22.73%), PrkE (13.64%)	CikA (12.12%), SasA (3.03%), LabA (1.51%), LalA (33.33%), Crm (7.58%), CpmA (63.64%)
Other	KaiB (31.43%), KaiC (100%)	CikA (25.71%), NhtA (17.14%), PrkE (71.43%), IrcA (54.29%)	CikA (25.71%), SasA (31.43%), LalA (5.71%), Crm (2.86%), RpaA (11.43%), RpaB (48.57%), CpmA (51.43%)

Table S3: **A collection of circadian and diurnal expression datasets in the cyanobacterial clade.** Datasets with assigned abbreviation were used to determine the core oscillatory genome. Information is provided about the ability to fix nitrogen, the habitat (freshwater F, saltwater S), the total number of genes, the publication reference, the absolute and relative number of diurnally expressed genes reported in the original publications, the applied light and sampling schema, the experimental culture conditions, and the methods for microarray normalisation and oscillating gene detection. References to datasets employed in the following comparison are shown bold.

Strain	Strain Abbrev.	N_2 Fixation	Habitat	Total Genes	Ref.	Dataset Abbrev.	Diurnal Genes	Light Conditions	Culture Conditions	Normalization / Oscillation Detection
Prochlorcoccus marinus MED4	ProMED4	-	M	1766	[133]	zinser09	1403 (79%)	LD (14:10, T_{samp} 2h)	Pro99 Medium, stirred, 24°C, Batch	RMA, Fourier Analysis
Synechocystis sp. PCC 6803	Syn6803	-	F	3628	[76]	leh13	1133 (31%)	LD (12:12, T_{samp} irreg.)	BG11 Medium, air bubbling, 30°C, Batch	LOS, Fourier Analysis
					[103]	beck14	(27%)	LD (12:12, T_{samp} 2h)	BG11 Medium, air bubbling, 30°C, Batch	LOS, Fourier Analysis
					[72]	-	1349 (37%)	LD (14:10, T_{samp} 2h)	BG11 Medium, stirred, 3% CO_2air bubbling, 27°C, Turbidostat	Standorf Microarray Database standard, ANOVA and correlation with light
					[73]	-	237 (9%)	LL (T_{samp} 4h)	BG11 Medium, air bubbling, stirred, 30°C, Batch with manual dilution	LOWESS, modified Cosiner
Synechococcus elongatus PCC 7942	Syc7942	-	F	2719	[113]	vijayan09	1748 (64%)	LL (T_{samp} 4h)	BG11 Medium, 1% air CO_2 bubbling, 30°C, Continuous Culture	Loess and Quantile, Fourier Analysis
					[101]	ito09	800 (29%)	LL (T_{samp} 2h)	BG11 Medium, 30°C, Continuous Culture	Replicate Mean Polishing, Correlation to Sine
Microcystis aeruginosa PCC 7806	Mic7806	-	F	6360	[134]	straub11	1344 (21%)	LD (12:12, T_{samp} irreg.)	BG11 Medium, 1% air CO_2 bubbling, 22°C, Batch	LOWESS, significant difference to CT0
Anabaena sp. PCC 7120	Ana7120	•	F	6222	[75]	kushige13	78 (1.25%)	LL (T_{samp} 4h)	BG11 + N Medium, 30°C, Contiuous Culture	Replicate Mean Polishing, Correlation to Sine
Cyanothece ATCC 51142	Cyn51142	•	M	5354	[114]	stoeckel08	1445 ($\approx$ 30%)	LD (12:12, T_{samp} 4h)	ASP2 Medium, 30°C, air bubbling, Batch	LOWESS, Correlation Network
					[115]	toepel08	1424 ($\approx$ 20%)	LD (12:12)/24h, LL T_{samp} 4h)	ASP2 Medium, 30°C, Airlift Bioreactor	LOWESS, Differential Expression
					[74]	-	1400 (27%)	LD (6:6, T_{samp} 2h)	ASP2 Medium, 30°C, Airlift Bioreactor	LOWESS, Correlation Network

Table S4: **Oligonucleatides used for cloning.**

Number	Primer Name	Primer Sequence	ORF
1	fw-kaiC1-Syn6714-BamHI	CTACGGATCCAACTCACCCA TCGTTAACG	D082_30580
2	rev-kaiC1-Syn6714-NotI	GAAGCGGCCGCCTACTCGAC GGTTTTATC	D082_30580
3	fw-kaiC1-Npun29133-BamHI	CTACGGATCCAGTCAAAACG AGCAAG	Npun_R2886
4	rev-kaiC1-Npun29133-NotI	CGAAGCGGCCGCTTAGGGTT CGGAAC	Npun_R2886
5	fw-kaiC1-Cy7424-BamHI	CATAGGATCCAATGAACCCA TTCCCAACG	PCC7424_0599
6	rev-kaiC1-Cy7424-NotI	CATTGCGGCCGCTTATTCAT CTAAAGTTTTATC	PCC7424_0599
7	fw-kaiC3-Cy7424-BamHI	CGAAGGATCCAATCAAGACA ACGAAC	PCC7424_3006
8	rev-kaiC3-Cy7424-NotI	CTGTGCGGCCGCCTAAGACC GTTCTTCAAAC	PCC7424_3006
9	fw-kaiC3-Mic7806-BamHI	CTACGGATCCACGCAAAATA ATCCCCTAG	IPF_2046
10	rev-kaiC3-Mic7806-NotI	GAAGCGGCCGCCTAACTACG ATCCTCA	IPF_2046
11	fw-KaiC3-PH_RS03935 BamHI (DSM 12428)	CCCGGATCCGATGCTCTTAA TTGTTGGAACTCC	Gene ID: 1443164
12	rev-KaiC3-PH_RS03935 HindIII (DSM 12428)	CGGGGAAGCTTTTACTCATA AATTTCCACCCCTC	Gene ID: 1443164
13	fw-KaiC3-OCC_RS02010 PstI (DSM 5473)	GGGCTGCAGATGAGCAGAAC GGGAATTG	Gene ID: 16548747
14	rev-KaiC3-OCC_RS02010 HindIII (DSM 5473)	GCCGGGAAGCTTTTATTCAT AAATTTCCACCC	Gene ID: 16548747

Table S5: **Diurnal core CLOGs across cyanobacterial strains** excluding *Microcystis aeruginosa* PCC 7806. In every CLOG (row) at least one gene of each of the considered datasets (columns) exhibited diurnal expression.

Syn6803	Syc7942	ProMED4	Cyn51142	Ana7120	Function
sll0158	1085	PMM0584	cce_2248	all0713	1,4-alpha-glucan branching enzyme
sll1817	2210	PMM1536	cce_4038	all4192	30S ribosomal protein S11
sll1096	0887	PMM1511	cce_4091	all4340	30S ribosomal protein S12
ssl3437	2223	PMM1549	cce_4023	asl4206	30S ribosomal protein S17
ssl3432	2228	PMM1554	cce_4018	asl4211	30S ribosomal protein S19
sll1260	2530	PMM0753	cce_0705	all4792	30S ribosomal protein S2
sll1804	2226	PMM1552	cce_4020	all4209	30S ribosomal protein S3
sll1812	2216	PMM1542	cce_4031	all4199	30S ribosomal protein S5
sll1097	0886	PMM1510	cce_4090	all4339	30S ribosomal protein S7
sll1809	2219	PMM1545	cce_4028	all4202	30S ribosomal protein S8
sll1822	2205	PMM1531	cce_4043	all4187	30S ribosomal protein S9
sll1821	2206	PMM1532	cce_4042	all4188	50S ribosomal protein L13
sll1802	2229	PMM1555	cce_4017	all4212	50S ribosomal protein L2
slr1678	1219	PMM1344	cce_1391	all0147	50S ribosomal protein L21
sll1803	2227	PMM1553	cce_4019	all4210	50S ribosomal protein L22
sll1807	2221	PMM1547	cce_4025	asl4204	50S ribosomal protein L24
sll1799	2232	PMM1558	cce_4013	all4215	50S ribosomal protein L3
sll1800	2231	PMM1557	cce_4015	all4214	50S ribosomal protein L4
sll1810	2218	PMM1544	cce_4029	all4201	50S ribosomal protein L6
sll0329	0039	PMM0770	cce_3746	alr5275	6-phosphogluconate dehydrogenase
sll1323	0333	PMM1454	cce_4485	all0008	ATP synthase subunit b' of CF(0)

sll1908	1501	PMM1354	cce_2134	alr1890	D-3-phosphoglycerate dehydrogenase
sll0519	1343	PMM0160	cce_2224	alr0223	NADH dehydrogenase subunit 1
slr0331	1976	PMM0150	cce_4299	alr3957	NADH dehydrogenase subunit 4 (involved in photosystem-1 cyclic electron flow)
sll0689	2359	PMM1600	cce_1688, cce_2468	alr0091, all1303	Na+/H+ antiporter
sll1818	2209	PMM1535	cce_4039	all4191	RNA polymerase alpha subunit
slr1265	1523	PMM1484	cce_3838	alr1595	RNA polymerase gamma-subunit
sll0306, sll2012	1746, 0672	PMM1629	cce_3594, cce_0644	alr3810, alr3800	RNA polymerase group 2 sigma factor
slr1424	1740	PMM0021	cce_2372	alr5066	UDP-N-acetylenolpyruvoylglucosamine reductase
sll0108	0442	PMM0263	cce_3261	alr0991	ammonium/methylammonium permease
ssl2667	0450	PMM0418	cce_1017	asr1309	an assembly factor for iron-sulfur culsters
slr0585	0009	PMM1707	cce_4370	alr4798	argininosuccinate synthetase
slr0549	1848	PMM1654	cce_0293	all3680	aspartate beta-semialdehyde dehydrogenese
slr1720	1313	PMM1688	cce_4152	all2436	aspartyl-tRNA synthetase
sll1498	2122	PMM0951	cce_0902	alr1155	carbamoyl-phosphate synthase small chain
sll1028, sll1029	1421	PMM0549	cce_4283, cce_4282	alr0867, all0868	carbon dioxide concentrating mechanism protein CcmK
slr1390	0942	PMM0743	cce_1270	all3642	cell division protein FtsH
sll0109	1915	PMM1181	cce_3372	all4012	chorismate mutase
slr0757	1217	PMM1343	cce_0423	alr2885	circadian clock protein KaiB homolog
slr1138	2604	PMM0444	cce_1975	alr0952	cytochrome c oxidase subunit III
slr0550	1847	PMM1653	cce_0294	all3679	dihydrodipicolinate synthase
slr1626	0040	PMM0591	cce_4420	alr3512	dihydroneopterin aldolase
slr2026	2303	PMM0830	cce_4305	alr4386	dihydropteroate synthase
slr1051	0126	PMM0282	cce_0460	all4391	enoyl-[acyl-carrier-protein] reductase
sll0018	1443	PMM0781	cce_1357	all4563	fructose-bisphosphate aldolase, class II
slr1843	2334	PMM1074	cce_2536	all4019	glucose 6-phosphate dehydrogenase
slr0638	2457	PMM1165	cce_3990	all1985	glycyl-tRNA synthetase alpha chain
slr1718	2589	PMM0614	cce_1018	all2568	hypothetical protein
sll0098	1758	PMM1481	cce_2727	all0355	hypothetical protein
sll0996	2374	PMM1305	cce_2810	alr1312	hypothetical protein
slr1847	0464	PMM0020	cce_4379	alr5067	hypothetical protein
slr1471	1617	PMM1186	cce_1364	alr3415	hypothetical protein
slr1577	0426	PMM0909	cce_0429	all5166	hypothetical protein
ssr0332	0658	PMM1563	cce_4619	asr4076	hypothetical protein
slr0742	2023	PMM1491	cce_1842	alr3828	hypothetical protein
ssl0353	2017	PMM0061	cce_0650	asl0940	hypothetical protein
slr0362	2396	PMM0789	cce_0268	alr3102	hypothetical protein
sll1898	2601	PMM0447	cce_4599	all0949	hypothetical protein
slr1896	0286	PMM0390	cce_2055	all0876	hypothetical protein
sll0372	0362	PMM0239	cce_2501	all2849	hypothetical protein
sll1866	1864	PMM0319	cce_4203	alr0116	hypothetical protein
sll0854	0826	PMM1517	cce_3387	all3378	hypothetical protein
slr0506	2503	PMM0542	cce_0320	all1743	light-dependent NADPH-protochlorophyllide oxidoreductase
slr0772	1838	PMM0544	cce_1954	alr3441	light-independent protochlorophyllide reductase subunit ChlB
slr1540	1826	PMM0797	cce_3779	alr4831	mRNA-binding protein

slr1055	2137	PMM0831	cce_4358	all4365	magnesium protoporphyrin IX chelatase subunit H
slr2033	1179	PMM0295	cce_1309	alr3843	membrane-associated rubredoxin, essential for photosystem I assembly
sll0902	2514	PMM1263	cce_3251	alr4907	ornithine carbamoyltransferase
sll1553	1293	PMM0871	cce_1321	alr4958	phenylalanyl-tRNA synthetase
sll0684	2441	PMM0725	cce_0883	all4572	phosphate transport ATP-binding protein PstB homolog
slr0597	0396	PMM0266	cce_4354	all3093	phosphoribosyl aminoimidazole carboxy formyl formyltransferase/inosinemonophosphate cyclohydrolase (PUR-H(J))
slr1645	0343	PMM0507	cce_3633	all1258	photosystem II 11 kD protein
sll0851	0656	PMM1158	cce_0659	alr4291	photosystem II CP43 protein
slr1311, sll1867, slr1181	0893, 1389, 0424	PMM0223	cce_3411, cce_0267, cce_3501, cce_0636	alr3742, alr4866, alr4592, alr3727, all3572	photosystem II D1 protein
slr0906	0697	PMM0315	cce_1837	all0138	photosystem II core light harvesting protein
sll0427	0294	PMM0228	cce_2572	all3854	photosystem II manganese-stabilizing polypeptide
sll0849, slr0927	1637, 0655	PMM1157	cce_2485, cce_0660	alr4290, alr4548	photosystem II reaction center D2 protein
sll0171	2308	PMM1687	cce_4346	all4609	probable aminomethyltransferase
slr1673	1199	PMM1299	cce_0402	alr0175	probable tRNA/rRNA methyltransferase
slr0774	0142	PMM0929	cce_4646	all0121	protein-export membrane protein SecD
sll1786	1521	PMM1486	cce_3489	alr1593	putative deoxyribonuclease, tatD homolog
slr2034	1178	PMM0296	cce_1308	alr3844	putative homolog of plant HCF136, which is essential for stability or assembly of photosystem II
sll1841	1068	PMM0405	cce_2750	alr3606	pyruvate dehydrogenase dihydrolipoamide acetyltransferase component (E2)
sll1282	2244	PMM1643	cce_4679	alr3993	riboflavin synthase beta subunit
slr0194	0584	PMM1489	cce_0103	all0888	ribose 5-phosphate isomerase
slr0012	1427	PMM0551	cce_3164	alr1526	ribulose bisphosphate carboxylase small subunit
slr0743	2022	PMM1492	cce_1841	alr3829	similar to N utilization substance protein
ssr1600	2121	PMM0950	cce_1801	asr1156	similar to anti-sigma f factor antagonist
sll1820	2207	PMM1533	cce_4041	all4189	tRNA pseudouridine synthase 1
sll1980	2128	PMM0242	cce_0972	alr0570	thiol:disulfide interchange protein TrxA
sll1615	1582	PMM0189	cce_4596	all4677	thiophen and furan oxidation protein
sll0755	2309	PMM0856	cce_2409	alr4641	thioredoxin peroxidase
slr1793	2297	PMM0519	cce_4687	all2563	transaldolase
sll1957	0688	PMM0714	cce_3556	alr2766	transcriptional regulator
slr1884	1308	PMM0598	cce_3534	all1269	tryptophanyl-tRNA synthetase

References

[1] Y. Ouyang, C. R. Andersson, T. Kondo, S. S. Golden, and C. H. Johnson. Resonating circadian clocks enhance fitness in cyanobacteria. *Proceedings of the National Academy of Sciences of the United States of America*, 95(15):8660–4, 1998.

[2] M. A. Woelfle, Y. Ouyang, K. Phanvijhitsiri, and C. H. Johnson. The adaptive value of circadian clocks: an experimental assessment in cyanobacteria. *Current Biology*, 14(16):1481–6, 2004.

[3] M. J. Simons. The evolution of the cyanobacterial posttranslational clock from a primitive "phoscillator". *Journal of Biological Rhythms*, 24(3):175–82, 2009.

[4] J.L. Ditty, S.B. Williams, and S.S. Golden. A Cyanobacterial Circadian Timing Mechanism. *Annual Review of Genetics*, 37(1):513–543, 2003.

[5] C. H. Johnson, P. L. Stewart, and M. Egli. The cyanobacterial circadian system: from biophysics to bioevolution. *Annual Review of Biophysics*, 40:143–67, 2011.

[6] S. S. Golden and S. R. Canales. Cyanobacterial circadian clocks–timing is everything. *Nature Review Microbiology*, 1(3):191–9, 2003.

[7] C. S. Pittendrigh. Temporal organization: reflections of a darwinian clock-watcher. *Annual Review of Physiology*, 55:16–54, 1993.

[8] B. M. Sweeney and M. B. Borgese. A circadian rhythm in cell division in a prokaryote, the cyanobacterium synechococcus wh7803. *Journal of Phycology*, 25(1):183–186, 1989.

[9] A. Mitsui, S. Kumazawa, A. Takahashi, H. Ikemoto, S. Cao, and T. Arai. Strategy by which nitrogen-fixing unicellular cyanobacteria grow photoautotrophically. *Nature*, 323(6090):720–722, 1986.

[10] T. C. Huang and T. J. Chow. New type of n-2-fixing unicellular cyanobacterium (blue-green-alga). *FEMS Microbiology Letters*, 36(1):109–110, 1986.

[11] N. Grobbelaar, T. C. Huang, H. Y. Lin, and T. J. Chow. Dinitrogen-fixing endogenous rhythm in synechococcus rf-1. *FEMS Microbiology Letters*, 37(2):173–177, 1986.

[12] T. C. Huang and T. J. Chow. Characterization of the rhythmic nitrogen-fixing activity of synechococcus sp rf-1 at the transcription level. *Current Microbiology*, 20(1):23–26, 1990.

[13] J. Tomita, M. Nakajima, T. Kondo, and H. Iwasaki. No transcription-translation feedback in circadian rhythm of kaic phosphorylation. *Science*, 307(5707):251–4, 2005.

[14] M. Nakajima, K. Imai, H. Ito, T. Nishiwaki, Y. Murayama, H. Iwasaki, T. Oyama, and T. Kondo. Reconstitution of circadian oscillation of cyanobacterial kaic phosphorylation in vitro. *Science*, 308(5720):414–5, 2005.

[15] J. L. Ditty, S. R. Canales, B. E. Anderson, S. B. Williams, and S. S. Golden. Stability of the synechococcus elongatus pcc 7942 circadian clock under directed anti-phase expression of the kai genes. *Microbiology*, 151(Pt 8):2605–2613, 2005.

[16] R. S. Edgar, E. W. Green, Y. Zhao, G. van Ooijen, M. Olmedo, X. Qin, Y. Xu, M. Pan, U. K. Valekunja, K. A. Feeney, E. S. Maywood, M. H. Hastings, N. S. Baliga, M. Merrow, A. J. Millar, C. H. Johnson, C. P. Kyriacou, J. S. O'Neill, and A. B. Reddy. Peroxiredoxins are conserved markers of circadian rhythms. *Nature*, 485(7399):459–64, 2012.

[17] J. S. O'Neill and A. B. Reddy. Circadian clocks in human red blood cells. *Nature*, 469(7331):498–503, 2011.

[18] J. S. O'Neill, G. van Ooijen, L. E. Dixon, C. Troein, F. Corellou, F.-Y. Bouget, A. B. Reddy, and A. J. Millar. Circadian rhythms persist without transcription in a eukaryote. *Nature*, 469(7331):554–8, 2011.

[19] C. R. McClung. Circadian rhythms in plants. *Annual Review of Plant Physiology and Plant Molecular Biology*, 52:139–162, 2001.

[20] Natalia B Ivleva, Matthew R Bramlett, Paul a Lindahl, and Susan S Golden. Ldpa: a component of the circadian clock senses redox state of the cell. *The EMBO Journal*, 24(6):1202–1210, 2005.

[21] Gopal K. Pattanayak and Michael J. Rust. The Cyanobacterial Clock and Metabolism. *Current Opinion in Microbiology*, 18:90–95, 2014.

[22] Y.-I. Kim, D. J. Vinyard, G. M. Ananyev, G. C. Dismukes, and S. S. Golden. Oxidized quinones signal onset of darkness directly to the cyanobacterial circadian oscillator. *Proceedings of the National Academy of Sciences of the United States of America*, 109(44):17765–17769, 2012.

[23] Y. Hihara, K. Sonoike, M. Kanehisa, and M. Ikeuchi. Dna microarray analysis of redox-responsive genes in the genome of the cyanobacterium synechocystis sp. strain pcc 6803. *Journal of Bacteriology*, 185(5):1719–25, 2003.

[24] F. Hayashi, H. Suzuki, R. Iwase, T. Uzumaki, A. Miyake, J. R. Shen, K. Imada, Y. Furukawa, K. Yonekura, K. Namba, and M. Ishiura. Atp-induced hexameric ring structure of the cyanobacterial circadian clock protein kaic. *Genes to Cells*, 8(3):287–96, 2003.

[25] T. Mori, S. V. Saveliev, Y. Xu, W. F. Stafford, M. M. Cox, R. B. Inman, and C. H. Johnson. Circadian clock protein kaic forms atp-dependent hexameric rings and binds dna. *Proceedings of the National Academy of Sciences of the United States of America*, 99(26):17203–8, 2002.

[26] T Nishiwaki, H Iwasaki, M Ishiura, and T Kondo. Nucleotide binding and autophosphorylation of the clock protein KaiC as a circadian timing process of cyanobacteria. *Proceedings of the National Academy of Sciences of the United States of America*, 97(1):495–499, 2000.

[27] Taeko Nishiwaki and Takao Kondo. Circadian autodephosphorylation of cyanobacterial clock protein KaiC occurs via formation of ATP as intermediate. *Journal of Biological Chemistry*, 287(22):18030–18035, 2012.

[28] Y.-I. Kim, G. Dong, Jr. Carruthers, C. W., S. S. Golden, and A. LiWang. The day/night switch in kaic, a central oscillator component of the circadian clock of cyanobacteria. *Proceedings of the National Academy of Sciences of the United States of America*, 105(35):12825–30, 2008.

[29] R. Pattanayek, D. R. Williams, S. Pattanayek, Y. Xu, T. Mori, C. H. Johnson, P. L. Stewart, and M. Egli. Analysis of kaia-kaic protein interactions in the cyano-bacterial circadian clock using hybrid structural methods. *The EMBO Journal*, 25(9):2017–28, 2006.

[30] Michael J Rust, Susan S Golden, and Erin K O'Shea. Light-driven changes in energy metabolism directly entrain the cyanobacterial circadian oscillator. *Science*, 331(6014):220–3, 2011.

[31] Y. Kitayama, H. Iwasaki, T. Nishiwaki, and T. Kondo. Kaib functions as an attenuator of kaic phosphorylation in the cyanobacterial circadian clock system. *The EMBO Journal*, 22(9):2127–34, 2003.

[32] S. B. Williams, I. Vakonakis, S. S. Golden, and A. C. LiWang. Structure and function from the circadian clock protein kaia of synechococcus elongatus: a potential clock input mechanism. *Proceedings of the National Academy of Sciences of the United States of America*, 99(24):15357–62, 2002.

[33] T. L. Wood, J. Bridwell-Rabb, Y. I. Kim, T. Gao, Y. G. Chang, A. LiWang, D. P. Barondeau, and S. S. Golden. The kaia protein of the cyanobacterial circadian oscillator is modulated by a redox-active cofactor. *Proceedings of the National Academy of Sciences of the United States of America*, 107(13):5804–9, 2010.

[34] Shinsuke Kutsuna, Takao Kondo, and Setsuyuki Aoki. A Period-Extender Gene , pex , That Extends the Period of the Circadian Clock in the Cyanobacterium Synechococcus sp . Strain A Period-Extender Gene , pex , That Extends the Period of the Circadian Clock in the Cyanobacterium Synechococcus sp . Strain PCC. *Journal of Bacteriology*, 180(8):2167–2174, 1998.

[35] Shinsuke Kutsuna, Takao Kondo, Haruki Ikegami, Tatsuya Uzumaki, Mitsunori Katayama, and Masahiro Ishiura. The circadian clock-related gene pex regulates a negative cis element in the kaiA promoter region. *Journal of Bacteriology*, 189(21):7690–7696, 2007.

[36] Natalia B Ivleva, Tiyu Gao, Andy C LiWang, and Susan S Golden. Quinone sensing by the circadian input kinase of the cyanobacterial circadian clock. *Proceedings of the National Academy of Sciences of the United States of America*, 103(46):17468–17473, 2006.

[37] O Schmitz, M Katayama, S B Williams, T Kondo, and S S Golden. Cika, a bacteriophytochrome that resets the cyanobacterial circadian clock. *Science*, 289(5480):765–768, 2000.

[38] Shannon R. Mackey, Jong Soon Choi, Yohko Kitayama, Hideo Iwasaki, Guogang Dong, and Susan S. Golden. Proteins found in a CikA interaction assay link the orcadian clock, metabolism, and cell division in synechococcus elongatus. *Journal of Bacteriology*, 190(10):3738–3746, 2008.

[39] Hideo Iwasaki, Stanly B. Williams, Yohko Kitayama, Masahiro Ishiura, Susan S. Golden, and Takao Kondo. A KaiC-interacting sensory histidine kinase, SasA, necessary to sustain robust circadian oscillation in cyanobacteria. *Cell*, 101(2):223–233, 2000.

[40] R. M. Smith and S. B. Williams. Circadian rhythms in gene transcription imparted by chromosome compaction in the cyanobacterium synechococcus elongatus. *Proceedings of the National Academy of Sciences of the United States of America*, 103(22):8564–9, 2006.

[41] Naoki Takai, Masato Nakajima, Tokitaka Oyama, Ryotaku Kito, Chieko Sugita, Mamoru Sugita, Takao Kondo, and Hideo Iwasaki. A KaiC-associating SasA-RpaA two-component regulatory system as a major circadian timing mediator in cyanobacteria. *Proceedings of the National Academy of Sciences of the United States of America*, 103(32):12109–12114, 2006.

[42] Mitsumasa Hanaoka, Naoki Takai, Norimune Hosokawa, Masayuki Fujiwara, Yuki Akimoto, Nami Kobori, Hideo Iwasaki, Takao Kondo, and Kan Tanaka. RpaB, another response regulator operating circadian clock-dependent transcriptional regulation in Synechococcus elongatus PCC 7942. *Journal of Biological Chemistry*, 287(31):26321–26327, 2012.

[43] Joseph S. Markson, Joseph R. Piechura, Anna M. Puszynska, and Erin K. O'Shea. Circadian control of global gene expression by the cyanobacterial master regulator rpaa. *Cell*, 155(6):1396–1408, 2013.

[44] Andrian Gutu and Erin K. O'Shea. Two Antagonistic Clock-Regulated Histidine Kinases Time the Activation of Circadian Gene Expression. *Molecular Cell*, 50(2):288–294, 2013.

[45] Yasuhito Taniguchi, Tomoe Nishikawa, Takao Kondo, and Tokitaka Oyama. Overexpression of lalA, a paralog of labA, is capable of affecting both circadian gene expression and cell growth in the cyanobacterium Synechococcus elongatus PCC 7942. *FEBS Letters*, 586(6):753–759, 2012.

[46] Joseph S Boyd, Juliana R Bordowitz, Anna C Bree, and Susan S Golden. An allele of the crm gene blocks cyanobacterial circadian rhythms. *Proceedings of the National Academy of Sciences of the United States of America*, 110(34):13950–5, 2013.

[47] Javier Espinosa, Joseph S. Boyd, Raquel Cantos, Paloma Salinas, Susan S. Golden, and Asuncion Contreras. Cross-talk and regulatory interactions between the essential response regulator RpaB and cyanobacterial circadian clock output. *Proceedings of the National Academy of Sciences of the United States of America*, page 201424632, 2015.

[48] F. Moronta-Barrios, J. Espinosa, and A. Contreras. In vivo features of signal transduction by the essential response regulator rpab from synechococcus elongatus pcc 7942. *Microbiology*, 158(Pt 5):1229–37, 2012.

[49] Mitsunori Katayama, Nicholas F. Tsinoremas, Takao Kondo, and Susan S. Golden. cpma, a gene involved in an output pathway of the cyanobacterial circadian system. *Journal of Bacteriology*, 181(11):3516–3524, 1999.

[50] Yasuhito Taniguchi, Mitsunori Katayama, Rie Ito, Naoki Takai, Takao Kondo, and Tokitaka Oyama. laba : a novel gene required for negative feedback regulation of the cyanobacterial circadian clock protein kaic. *Genes and Development*, pages 60–70, 2007.

[51] F. Moronta-Barrios, J. Espinosa, and A. Contreras. Negative control of cell size in the cyanobacterium synechococcus elongatus pcc 7942 by the essential response regulator rpab. *FEBS Letters*, 587(5):504–9, 2013.

[52] V. Dvornyk. *The Circadian Clock Gear in Cyanobacteria: Assembled by Evolution*, book section 14, pages 241–258. Springer, Berlin Heidelberg, 2009.

[53] Volodymyr Dvornyk, Oxana Vinogradova, and Eviatar Nevo. Origin and evolution of circadian clock genes in prokaryotes. *Proceedings of the National Academy of Sciences of the United States of America*, 100(5):2495–2500, 2003.

[54] Ilka M. Axmann, Stefanie Hertel, Anika Wiegard, Anja K. Dörrich, and Annegret Wilde. Diversity of KaiC-based timing systems in marine Cyanobacteria. *Marine Genomics*, 14:3–16, 2014.

[55] Anika Wiegard, Anja K. Dörrich, Hans Tobias Deinzer, Christian Beck, Annegret Wilde, Julia Holtzendorff, and Ilka M. Axmann. Biochemical analysis of three putative kaic clock proteins from synechocystis sp. pcc 6803 suggests their functional divergence. *Microbiology*, 159(PART 5):948–958, 2013.

[56] S. Aoki and K. Onai. *Circadian clocks of Synechocystis sp. strain PCC 6803, Thermosynechococcus elongatus, Prochlorococcus spp., Trichodesmium spp. and other species*. Springer Berlin Heidelberg, 2009.

[57] P. Ma, T. Mori, C. Zhao, T. Thiel, and C. H. Johnson. Evolution of kaic-dependent timekeepers: A proto-circadian timing mechanism confers adaptive fitness in the purple bacterium rhodopseudomonas palustris. *PLoS Genetics*, 12(3):e1005922, 2016.

[58] H. J. Kang, K. Kubota, H. Ming, K. Miyazono, and M. Tanokura. Crystal structure of kaic-like protein ph0186 from hyperthermophilic archaea pyrococcus horikoshii ot3. *Proteins*, 75(4):1035–9, 2009.

[59] M. Maniscalco, J. Nannen, V. Sodi, G. Silver, P. L. Lowrey, and K. A. Bidle. Light-dependent expression of four cryptic archaeal circadian gene homologs. *Frontiers in Microbiology*, 5:79, 2014.

[60] Ivan Baca, Daniel Sprockett, and Volodymyr Dvornyk. Circadian input kinases and their homologs in cyanobacteria: Evolutionary constraints versus architectural diversification. *Journal of Molecular Evolution*, 70(5):453–465, 2010.

[61] Volodymyr Dvornyk, Hong Wen Deng, and Eviatar Nevo. Structure and molecular phylogeny of sasa genes in cyanobacteria: Insights into evolution of the prokaryotic circadian system. *Molecular Biology and Evolution*, 21(8):1468–1476, 2004.

[62] P. J. Cock, T. Antao, J. T. Chang, B. A. Chapman, C. J. Cox, A. Dalke, I. Friedberg, T. Hamelryck, F. Kauff, B. Wilczynski, and M. J. de Hoon. Biopython: freely available python tools for computational molecular biology and bioinformatics. *Bioinformatics*, 25(11):1422–3, 2009.

[63] Fernando Pérez and E. Granger. Ipython: A system for interactive scientific computing, computing in science and engineering. *Computing in Science and Engineering*, 9(3):21–29, 2007.

[64] Stefan van der Walt, S. Chris Colbert, and Gael Varoquaux. The numpy array: A structure for efficient numerical computation. *Computing in Science and Engineering*, 13(2):22–30, 2011.

[65] J. D. Hunter. Matplotlib: A 2d graphics environment. *Computing in Science and Engineering*, 9(3):90–95, 2007.

[66] Wes McKinney. Data structures for statistical computing in python. In *Proceedings of the 9th Python in Science Conference*, volume 445, pages 51–56.

[67] D. A. Benson, I. Karsch-Mizrachi, D. J. Lipman, J. Ostell, and D. L. Wheeler. Genbank. *Nucleic Acids Research*, 36(Database issue):D25–30, 2008.

[68] Mitsuteru Nakao, Shinobu Okamoto, Mitsuyo Kohara, Tsunakazu Fujishiro, Takatomo Fujisawa, Shusei Sato, Satoshi Tabata, Takakazu Kaneko, and Yasukazu Nakamura. Cyanobase: the cyanobacteria genome database update 2010. *Nucleic Acids Research*, 38(Database issue):D379–D381, 2010.

[69] S. F. Altschul, T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. Gapped blast and psi-blast: a new generation of protein database search programs. *Nucleic Acids Research*, 25(17):3389–402, 1997.

[70] R. A. Fisher. On the interpretation of χ^2 from contingency tables, and the calculation of p. *Journal of the Royal Statistical Society*, 85(1):87–94, 1922.

[71] Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1):289–300, 1995.

[72] Rochelle G. Labiosa, Kevin R. Arrigo, Chao J. Tu, Devaki Bhaya, Stephen Bay, Arthur R. Grossman, and Jeff Shrager. Examination of diel changes in global transcript accumulation in synechocystis (cyanobacteria). *Journal of Phycology*, 42(3):622–636, 2006.

[73] K.-i. Kucho, K. Okamoto, Y. Tsuchiya, S. Nomura, M. Nango, M. Kanehisa, and M. Ishiura. Global analysis of circadian expression in the cyanobacterium synechocystis sp. strain pcc 6803. *Journal of Bacteriology*, 187(6):2190–9, 2005.

[74] J. R. Toepel, J. E. McDermott, T. C. Summerfield, and L. A. Sherman. Transcriptional analysis of the unicellular, diazotrophic cyanobacterium cyanothece sp. 51142 grown under short day/night cycles. *Journal of Phycology*, 45(3):610–20, 2009.

[75] H. Kushige, H. Kugenuma, M. Matsuoka, S. Ehira, M. Ohmori, and H. Iwasaki. Genome-wide and heterocyst-specific circadian gene expression in the filamentous cyanobacterium anabaena sp. strain pcc 7120. *Journal of Bacteriology*, 195(6):1276–84, 2013.

[76] Robert Lehmann, Rainer Machné, Jens Georg, Manuela Benary, Ilka Axmann, and Ralf Steuer. How cyanobacteria pose new problems to old methods: challenges in microarray time series analysis. *BMC Bioinformatics*, 14:133, 2013.

[77] Qiong Yang, Bernardo F. Pando, Guogang Dong, Susan S. Golden, and Alexander van Oudenaarden. Circadian gating of the cell cycle revealed in single cyanobacterial cells. *Science*, 327(5972):1522–1526, 2010.

[78] P. O. Westermark and H. Herzel. Mechanism for 12 hr rhythm generation by the circadian clock. *Cell Reports*, 3(4):1228–38, 2013.

[79] F. Sievers, A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Soding, J. D. Thompson, and D. G. Higgins. Fast, scalable generation of high-quality protein multiple sequence alignments using clustal omega. *Molecular Systems Biology*, 7:539, 2011.

[80] Andrew M. Waterhouse, James B. Procter, David M. A. Martin, Michèle Clamp, and Geoffrey J. Barton. Jalview version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, 25(9):1189–1191, 2009.

[81] Gavin E. Crooks, Gary Hon, John-Marc Chandonia, and Steven E. Brenner. Weblogo: A sequence logo generator. *Genome Research*, 14(6):1188–1190, 2004.

[82] U. K. Laemmli. Cleavage of structural proteins during the assembly of the head of bacteriophage t4. *Nature*, 227(5259):680–5, 1970.

[83] Ilka M. Axmann, Ulf Dühring, Luiza Seeliger, Anne Arnold, Jens T. Vanselow, Achim Kramer, and Annegret Wilde. Biochemical evidence for a timing mechanism in Prochlorococcus. *Journal of Bacteriology*, 191(17):5342–5347, 2009.

[84] M. Loza-Correa, L. Gomez-Valero, and C. Buchrieser. Circadian clock proteins in prokaryotes: hidden rhythms? *Frontiers in Microbiology*, 1:130, 2010.

[85] Takuro Nakayama, Ryoma Kamikawa, Goro Tanifuji, Yuichiro Kashiyama, Naohiko Ohkouchi, John M. Archibald, and Yuji Inagaki. Complete genome of a nonphotosynthetic cyanobacterium in a diatom reveals recent adaptations to an intracellular lifestyle. *Proceedings of the National Academy of Sciences of the United States of America*, 111(31):11407–11412, 2014.

[86] Yasukazu Nakamura, Takakazu Kaneko, Shusei Sato, Mamoru Mimuro, Hideaki Miyashita, Tohru Tsuchiya, Shigemi Sasamoto, Akiko Watanabe, Kumiko Kawashima, Yoshie Kishida, Chiaki Kiyokawa, Mitsuyo Kohara, Midori Matsumoto, Ai Matsuno, Naomi Nakazaki, Sayaka Shimpo, Chie Takeuchi, Manabu Yamada, and Satoshi Tabata. Complete genome structure of gloeobacter violaceus pcc 7421, a cyanobacterium that lacks thylakoids. *DNA Research*, 10(4):137–145, 2003.

[87] Julia Holtzendorff, Frédéric Partensky, Daniella Mella, Jean-François Lennon, Wolfgang R Hess, and Laurence Garczarek. Genome streamlining results in loss of robustness of the circadian clock in the marine cyanobacterium Prochlorococcus marinus PCC 9511. *Journal of Biological Rhythms*, 23(3):187–199, 2008.

[88] Ryan K Shultzaberger, Joseph S Boyd, Takeo Katsuki, Susan S Golden, and Ralph J Greenspan. Single mutations in sasa enable a simpler δcika gene network architecture with equivalent circadian properties. *Proceedings of the National Academy of Sciences of the United States of America*, 111(47):E5069–75, 2014.

[89] Volodymyr Dvornyk. Molecular evolution of ldpA, a gene mediating the circadian input signal in cyanobacteria. *Journal of Molecular Evolution*, 60(1):105–112, 2005.

[90] Mitsunori Katayama, Takao Kondo, Jin Xiong, S Susan, and Susan S Golden. ldpA Encodes an Iron-Sulfur Protein Involved in Light-Dependent Modulation of the Circadian Period in the Cyanobacterium Synechococcus elongatus PCC 7942. *Journal of Bacteriology*, 185(4):1415–1422, 2003.

[91] H. Min, H. Guo, and J. Xiong. Rhythmic gene expression in a purple photosynthetic bacterium, rhodobacter sphaeroides. *FEBS Letters*, 579(3):808–12, 2005.

[92] T. Nishiwaki-Ohkawa, Y. Kitayama, E. Ochiai, and T. Kondo. Exchange of adp with atp in the cii atpase domain promotes autophosphorylation of cyanobacterial clock protein kaic. *Proceedings of the National Academy of Sciences of the United States of America*, 111(12):4455–60, 2014.

[93] G. Dong, Q. Yang, Q. Wang, Y. I. Kim, T. L. Wood, K. W. Osteryoung, A. van Oudenaarden, and S. S. Golden. Elevated atpase activity of kaic applies a circadian checkpoint on cell division in synechococcus elongatus. *Cell*, 140(4):529–39, 2010.

[94] Y. Taniguchi, N. Takai, M. Katayama, T. Kondo, and T. Oyama. Three major output pathways from the kaiabc-based oscillator cooperate to generate robust circadian kaibc expression in cyanobacteria. *Proceedings of the National Academy of Sciences of the United States of America*, 2010.

[95] Ioannis Vakonakis, Douglas A. Klewer, Stanly B. Williams, Susan S. Golden, and Andy C. LiWang. Structure of the n-terminal domain of the circadian clock-associated histidine kinase sasa. *Journal of Molecular Biology*, 342(1):9–17, 2004.

[96] Yanmei Shi, Gene W Tyson, John M Eppley, and Edward F DeLong. Integrated metatranscriptomic and metagenomic analyses of stratified microbial assemblages in the open ocean. *The ISME Journal*, 5(6):999–1013, 2011.

[97] Annegret Wilde and Yukako Hihara. Transcriptional and posttranscriptional regulation of cyanobacterial photosynthesis. *Biochimica et Biophysica Acta - Bioenergetics*, 2015.

[98] Ryan K. Shultzaberger, Joseph S. Boyd, Spencer Diamonad, Ralph J. Greenspan, and Susan S. Golden. Giving time purpose: The synechococcus elongatus clock in a broader network context. *Annual Review of Genetics*, 49(1):485–505, 2015.

[99] Y. Liu, N. F. Tsinoremas, C. H. Johnson, N. V. Lebedeva, S. S. Golden, M. Ishiura, and T. Kondo. Circadian orchestration of gene expression in cyanobacteria. *Genes and Development*, 9(12):1469–78, 1995.

[100] Mark A. Woelfle and Carl Hirschie Johnson. No promoter left behind: Global circadian gene expression in cyanobacteria. *Journal of Biological Rhythms*, 21(6):419–431, 2006.

[101] H. Ito, M. Mutsuda, Y. Murayama, J. Tomita, N. Hosokawa, K. Terauchi, C. Sugita, M. Sugita, T. Kondo, and H. Iwasaki. Cyanobacterial daily life with kai-based circadian and diurnal genome-wide transcriptional control in synechococcus elongatus. *Proceedings of the National Academy of Sciences of the United States of America*, 106(33):14168–73, 2009.

[102] S Rao Jammalamadaka and Ashis SenGupta. *Topics in Circular Statistics*. World Scientific, 2001.

[103] Christian Beck, Stefanie Hertel, Anne Rediger, Robert Lehmann, Anika Wiegard, Adrian Kölsch, Beate Heilmann, Jens Georg, Wolfgang R. Hess, and Ilka M. Axmann. Daily expression pattern of protein-encoding genes and small noncoding RNAs in synechocystis sp. strain PCC 6803. *Applied and Environmental Microbiology*, 80(17):5195–5206, 2014.

[104] Otilia Cheregi, Cosmin Sicora, Peter B. Kos, Peter J. Nixon, and Imre Vass. The ftsh protease is required for the repair of photosystem ii in the cyanbacterium synechocystis6803 damaged uv-b radiation. *BMC Plant Biology*, 5(1):1–2, 2005.

[105] S. Jacobshagen, J. R. Whetstine, and J. M. Boling. Many but not all genes in chlamydomonas reinhardtii are regulated by the circadian clock. *Plant Biology*, 3(6):592–597, 2001.

[106] Stacey L. Harmer, John B. Hogenesch, Marty Straume, Hur-Song Chang, Bin Han, Tong Zhu, Xun Wang, Joel A. Kreps, and Steve A. Kay. Orchestrated transcription of key pathways in arabidopsis by the circadian clock. *Science*, 290(5499):2110–2113, 2000.

[107] Yukihiro Kabeya, Hiromitsu Nakanishi, Kenji Suzuki, Takanari Ichikawa, Youichi Kondou, Minami Matsui, and Shin-ya Miyagishima. The ylmg protein has a conserved function related to the distribution of nucleoids in chloroplasts and cyanobacteria. *BMC Plant Biology*, 10(1):1–13, 2010.

[108] K. Imai, T. Nishiwaki, T. Kondo, and H. Iwasaki. Circadian rhythms in the synthesis and degradation of a master clock protein kaic in cyanobacteria. *Journal of Biological Chemistry*, 279(35):36534–9, 2004.

[109] Hideo Iwasaki, Taeko Nishiwaki, Yohko Kitayama, Masato Nakajima, and Takao Kondo. KaiA-stimulated KaiC phosphorylation in circadian timing loops in cyanobacteria. *Proceedings of the National Academy of Sciences of the United States of America*, 99(24):15788–93, 2002.

[110] D. Zwicker, D. K. Lubensky, and P. R. Ten Wolde. Robust circadian clocks from coupled protein-modification and transcription-translation cycles. *Proceedings of the National Academy of Sciences of the United States of America*, 107(52):22540–5, 2010.

[111] Danish Memon, Abhay K. Singh, Himadri B. Pakrasi, and Pramod P. Wangikar. A global analysis of adaptive evolution of operons in cyanobacteria. *Antonie van Leeuwenhoek*, 103(2):331–346, 2013.

[112] M. Ishiura, S. Kutsuna, S. Aoki, H. Iwasaki, C. R. Andersson, A. Tanabe, S. S. Golden, C. H. Johnson, and T. Kondo. Expression of a gene cluster kaiabc as a circadian feedback process in cyanobacteria. *Science*, 281(5382):1519–23, 1998.

[113] Vikram Vijayan, Rick Zuzow, and Erin K O'Shea. Oscillations in supercoiling drive circadian gene expression in cyanobacteria. *Proceedings of the National Academy of Sciences of the United States of America*, 106(52):22564–8, 2009.

[114] J. Stöckel, Eric a Welsh, Michelle Liberton, Rangesh Kunnvakkam, Rajeev Aurora, and Himadri B. Pakrasi. Global transcriptomic analysis of cyanothece 51142 reveals robust diurnal oscillation of central metabolic processes. *Proceedings of the National Academy of Sciences of the United States of America*, 105(16):6156–6161, 2008.

[115] Jörg Toepel, Eric Welsh, Tina C. Summerfield, Himadri B. Pakrasi, and Louis A. Sherman. Differential transcriptional analysis of the cyanobacterium Cyanothece sp. strain ATCC 51142 during light-dark and continuous-light growth. *Journal of Bacteriology*, 190(11):3904–3913, 2008.

[116] Tatsuya Uzumaki, Masayasu Fujita, Toru Nakatsu, Fumio Hayashi, Hiroyuki Shibata, Noriyo Itoh, Hiroaki Kato, and Masahiro Ishiura. Crystal structure of the c-terminal clock-oscillator domain of the cyanobacterial kaia protein. *Nature Structural and Molecular Biology*, 11(7):623–631, 2004.

[117] H Iwasaki, Y Taniguchi, M Ishiura, and T Kondo. Physical interactions among circadian clock proteins KaiA, KaiB and KaiC in cyanobacteria. *The EMBO Journal*, 18(5):1137–1145, 1999.

[118] M. Egli, R. Pattanayek, J. H. Sheehan, Y. Xu, T. Mori, J. A. Smith, and C. H. Johnson. Loop-loop interactions regulate kaia-stimulated kaic phosphorylation in the cyanobacterial kaiabc circadian clock. *Biochemistry*, 52(7):1208–20, 2013.

[119] Y. Xu, T. Mori, R. Pattanayek, S. Pattanayek, M. Egli, and C. H. Johnson. Identification of key phosphorylation sites in the circadian clock protein kaic by crystallographic and mutagenetic analyses. *Proceedings of the National Academy of Sciences of the United States of America*, 101(38):13933–8, 2004.

[120] T. Nishiwaki, Y. Satomi, M. Nakajima, C. Lee, R. Kiyohara, H. Kageyama, Y. Kitayama, M. Temamoto, A. Yamaguchi, A. Hijikata, M. Go, H. Iwasaki, T. Takao, and T. Kondo. Role of kaic phosphorylation in the circadian clock system of synechococcus elongatus pcc 7942. *Proceedings of the National Academy of Sciences of the United States of America*, 101(38):13927–32, 2004.

[121] I. Vakonakis and A. C. LiWang. Structure of the c-terminal domain of the clock protein kaia in complex with a kaic-derived peptide: implications for kaic regulation. *Proceedings of the National Academy of Sciences of the United States of America*, 101(30):10925–30, 2004.

[122] Detlef D. Leipe, L. Aravind, Nick V. Grishin, and Eugene V. Koonin. The bacterial replicative helicase dnab evolved from a reca duplication. *Genome Research*, 10(1):5–16, 2000.

[123] Maria Loza-Correa, Tobias Sahr, Monica Rolando, Craig Daniels, Pierre Petit, Tania Skarina, Laura Gomez Valero, Delphine Dervins-Ravault, Nadine Honoré, Aleksey Savchenko, and Carmen Buchrieser. The Legionella pneumophila kai operon is implicated in stress response and confers fitness in competitive environments. *Environmental Microbiology*, 16(2):359–381, 2014.

[124] R. Pattanayek, J. Wang, T. Mori, Y. Xu, C. H. Johnson, and M. Egli. Visualizing a circadian clock protein: crystal structure of kaic and functional insights. *Molecular Cell*, 15(3):375–88, 2004.

[125] J. E. Walker, M. Saraste, M. J. Runswick, and N. J. Gay. Distantly related sequences in the alpha- and beta-subunits of atp synthase, myosin, kinases and other atp-requiring enzymes and a common nucleotide binding fold. *The EMBO Journal*, 1(8):945–51, 1982.

[126] M. Egli, T. Mori, R. Pattanayek, Y. Xu, X. Qin, and C. H. Johnson. Dephosphorylation of the core clock protein kaic in the cyanobacterial kaiabc circadian oscillator proceeds via an atp synthase mechanism. *Biochemistry*, 51(8):1547–58, 2012.

[127] Y. Kitayama, T. Nishiwaki-Ohkawa, Y. Sugisawa, and T. Kondo. Kaic intersubunit communication facilitates robustness of circadian rhythms in cyanobacteria. *Nature Communications*, 4:2897, 2013.

[128] R. Pattanayek, Y. Xu, A. Lamichhane, C. H. Johnson, and M. Egli. An arginine tetrad as mediator of input-dependent and input-independent atpases in the clock protein kaic. *Acta Crystallographica Section D: Biological Crystallography*, 70(Pt 5):1375–90, 2014.

[129] R. Pattanayek, T. Mori, Y. Xu, S. Pattanayek, C. H. Johnson, and M. Egli. Structures of kaic circadian clock mutant proteins: a new phosphorylation site at t426 and mechanisms of kinase, atpase and phosphatase. *PLoS One*, 4(11):e7529, 2009.

[130] Y. Xu, T. Mori, X. Qin, H. Yan, M. Egli, and C. H. Johnson. Intramolecular regulation of phosphorylation status of the circadian clock protein kaic. *PLoS One*, 4(11):e7509, 2009.

[131] Annemarie Neuner, Holger W. Jannasch, Shimshon Belkin, and Karl O. Stetter. Thermococcus litoralis sp. nov.: A new species of extremely thermophilic marine archaebacteria. *Archives of Microbiology*, 153(2):205–207, 1990.

[132] Juan M. Gonzalez, Yaeko Masuchi, Frank T. Robb, James W. Ammerman, Dennis L. Maeder, Miki Yanagibayashi, Jin Tamaoka, and Chiaki Kato. Pyrococcus horikoshii sp. nov., a hyperthermophilic archaeon isolated from a hydrothermal vent at the okinawa trough. *Extremophiles*, 2:123–130, 1998.

[133] Erik R. Zinser, Debbie Lindell, Zackary I. Johnson, Matthias E. Futschik, Claudia Steglich, Maureen L. Coleman, Matthew A. Wright, Trent Rector, Robert Steen, Nathan McNulty, Luke R. Thompson, and Sallie W. Chisholm. Choreography of the transcriptome, photophysiology, and cell cycle of a minimal photoautotroph, Prochlorococcus. *PLoS One*, 4(4), 2009.

[134] Cécile Straub, Philippe Quillardet, Julia Vergalli, Nicole Tandeau de Marsac, and Jean Francois Humbert. A day in the life of microcystis aeruginosa strain pcc 7806 as revealed by a transcriptomic analysis. *PLoS One*, 6(1), 2011.