A transcriptional cycle suited to daytime N_2 fixation in the unicellular cyanobacterium Candidatus *Atelocyanobacterium thalassa* (UCYN-A)

María del Carmen Muñoz-Marin[a, b], Irina N. Shilova[a#], Tuo Shi[a, c], Hanna Farnelid[a, d], Ana Maria Cabello[a] & Jonathan P. Zehr[a*]

[a] Ocean Sciences Department, University of California, Santa Cruz, CA, 95064, USA.

[b]Departamento de Bioquímica y Biología Molecular, Universidad de Córdoba, Córdoba, 14071, Spain.

[c]State Key Laboratory of Marine Environmental Science, Xiamen University, Xiamen, 361102, China.

[d]Centre for Ecology and Evolution in Microbial Model Systems (EEMiS), Linnaeus University, SE-391 82, Kalmar, Sweden.

Running Head: Diel whole genome transcription in UCYN-A

*Address correspondence to Prof. Jonathan Zehr, zehrj@ucsc.edu.

Present address: Second Genome, Inc, South San Francisco, CA, 94080, USA

Abstract

The symbiosis between a marine alga and a N_2-fixing cyanobacterium (UCYN-A) is geographically widespread in the oceans and is important in the marine N cycle. UCYN-A is uncultivated, and is an unusual unicellular cyanobacterium because it lacks many metabolic functions, including oxygenic photosynthesis and carbon fixation, which are typical in cyanobacteria. It is now presumed to be an obligate symbiont of haptophytes closely related to *Braarudosphaera bigelowii.* N_2-fixing cyanobacteria use different strategies to avoid inhibition of N_2 fixation by the oxygen evolved in photosynthesis. Most unicellular cyanobacteria temporally separate the two incompatible activities by fixing N_2 only at night, but surprisingly UCYN-A appears to fix N_2 during the day. The goal of this study was to determine how the unicellular UCYN-A coordinates N_2 fixation and general metabolism compared to other marine cyanobacteria. We found that UCYN-A has distinct daily cycles of many genes despite the fact that it lacks two of the three circadian clock genes found in most cyanobacteria. We also found that transcription patterns in UCYN-A are most similar to marine cyanobacteria that are capable of aerobic N_2 fixation in the light such as *Trichodesmium* and heterocyst-forming cyanobacteria, rather than *Crocosphaera* or *Cyanothece* species, which are more closely related to unicellular marine cyanobacteria evolutionarily. Our findings suggest that the symbiotic interaction has resulted in a shift of transcriptional regulation to coordinate UCYN-A metabolism with the phototrophic eukaryotic host, thus allowing efficient coupling of N_2 fixation (by the cyanobacterium) to the energy obtained from photosynthesis (by the eukaryotic unicellular alga) in the light.

Importance

The symbiotic N_2-fixing cyanobacterium UCYN-A and its eukaryotic algal host, which is closely related to *Braarudosphaera bigelowii*, have been shown to be globally distributed and important in open ocean N_2 fixation. These unique cyanobacteria have reduced metabolic capabilities, even lacking genes for oxygenic photosynthesis and carbon fixation. Cyanobacteria generally use energy from photosynthesis for nitrogen fixation, but require mechanisms for avoiding inactivation of the oxygen-sensitive nitrogenase enzyme by ambient oxygen (O_2) or the O_2 evolved through photosynthesis. This study shows that the symbiosis between the N_2-fixing cyanobacterium UCYN-A and its eukaryotic algal host has led to adaptation of its daily gene expression pattern in order to enable daytime aerobic N_2 fixation, which is likely more energetically efficient than fixing N_2 at night, as in other unicellular marine cyanobacteria.

Introduction

Nitrogen (N_2)-fixing microorganisms (diazotrophs), which reduce atmospheric N_2 to biologically available ammonium, are critical components of aquatic and terrestrial ecosystems because they supply fixed inorganic N (1). Cyanobacteria are particularly important in N_2 fixation because they can fuel the energy intensive N_2 reduction reaction using energy supplied by oxygenic photosynthesis. In the oceans, the filamentous, non-heterocyst-forming cyanobacterium *Trichodesmium* and the heterocyst-forming symbiont of diatoms (*Richelia* and related cyanobacteria) were believed to be the major N_2-fixing microorganisms until the discovery of the unicellular cyanobacteria *Crocosphaera*, *Cyanothece* and Candidatus *Atelocyanobacterium thalassa* (UCYN-A) in the open ocean. *Crocosphaera* and *Cyanothece* are free-living marine cyanobacteria, but UCYN-A is unusual in that it lacks oxygenic photosynthesis and is a symbiont of a haptophyte alga (related to *Braarudosphaera bigelowii*). The UCYN-A symbiosis is geographically widespread and is important in oceanic N_2 fixation (2-5). The UCYN-A genome has been greatly reduced, with massive metabolic streamlining including the loss of the oxygen-evolving Photosystem II (PSII), the carbon-fixing enzyme RuBisCO, and the entire tricarboxylic acid (TCA) cycle (6). UCYN-A has been shown to supply fixed N to the haptophyte in exchange for fixed carbon (4, 7), but it is not known how these two single-celled organisms coordinate metabolism and cell growth over the daily division cycle.

N_2 fixation requires energy and reductant, but the nitrogenase enzyme is inactivated by oxygen (O_2). Cyanobacteria generally have access to sufficient energy from photosynthesis but require mechanisms for avoiding inactivation of nitrogenase and N_2 fixation by ambient oxygen (O_2) or the O_2 evolved through photosynthesis. *Trichodesmium* and heterocystous cyanobacteria such as *Richelia* and *Nostoc* fix N_2 during the day, whereas the free-living unicellular *Crocosphaera* and *Cyanothece* fix N_2 at night. Interestingly, the symbiotic UCYN-A appears to fix

N_2 during the day (8-10), in contrast to most other unicellular marine N_2-fixing cyanobacteria, such as *Crocosphaera* and *Cyanothece*.

The processes of N_2-fixation and photosynthesis in cyanobacteria are regulated daily to increase cellular fitness and ecological competitiveness (11-13). Most cyanobacteria have circadian rhythms (11, 14, 15) that are involved in controlling daily cycles of gene transcription and protein synthesis by signal transduction pathways involving the circadian clock *kai* genes. UCYN-A lacks two of the three *kai* genes (*kaiA* and *kaiB*) known in most other cyanobacteria, whereas the non-N_2-fixing cyanobacterium *Prochlorococcus* only lacks *kaiA*. Thus, the daily whole genome expression pattern in UCYN-A is of interest to determine if there are daily patterns as in all other cyanobacteria compared to evolutionarily-related unicellular cyanobacteria.

We used a whole genome transcription array that targets two genetically distinct uncultivated sub-lineages of UCYN-A (UCYN-A1 and UCYN-A2), which have similar, but genetically distinct hosts. We compared the UCYN-A whole genome diel transcription patterns to those of *Cyanothece* sp. ATCC 51142 (16) and *Crocosphaera watsonii* WH 8501 (17) (both unicellular night-time N_2-fixers) and of *Trichodesmium erythraeum* IMS101 (a filamentous non-heterocystous day-time N_2-fixer). We also compared expression to whole genome expression of *Prochlorococcus* sp. MED4 (18) (a marine non-N_2-fixer) in order to determine how UCYN-A gene expression compares to general cyanobacterial gene expression in a sympatric open ocean species. We found that many genes in UCYN-A have distinct diel expression patterns and that UCYN-A has unusual gene expression patterns in comparison to unicellular N_2-fixing cyanobacteria that fix N_2 in the dark; however, it shares some general patterns with daytime N_2-fixing cyanobacteria, with heterocysts of heterocyst-forming cyanobacteria and with unicellular non-N_2-fixing cyanobacteria. Results suggest that optimal metabolism for open ocean cyanobacteria is aligned to the light period, and that symbiosis has enabled the unicellular UCYN-A to shift N_2 fixation to the daylight period.

Results and discussion

UCYN-A has a daily rhythm of gene transcription

UCYN-A has clear diel patterns of gene transcription, with a large fraction of genes that had periodicity of transcript levels over the dark and light periods (27%).

About a third of the UCYN-A genome (31% genes) targeted by the array were transcribed at detectable levels (365 of 1194 total genes in UCYN-A1 and 394 of 1244 total genes in UCYN-A2, respectively) (Table S1). Approximately 85% of these genes have differences in transcript levels between dark and light periods, accounting for 27% of the total genes in each strain (Table S1 and S2). *C. watsonii*, *Cyanothece* sp. and *Trichodesmium* cultures also had a large fraction of genes with changes in the transcript levels between dark and light periods (39% in *C. watsonii*, 20% in *Cyanothece* sp. and 34% in *Trichodesmium*) (Tables S1 and S2).

The UCYN-A transcription values (log_2-transformed) ranged from 2 to 13.5 with the median of 6.0. In both sub-lineages, genes coding for nitrogenase (*nif*), F_0F_1-ATP synthase (*atpA*, *atpB*), cytochrome b_6f complex (*petB*, *petC*, *petF*, *petL*) and the photosynthetic gene *psaC* were the most highly transcribed in comparison to all detected genes (Table S5). Transcript levels of the same genes were also high for both sub-lineages in metatranscriptomes collected during the TARA expedition in the South Atlantic Ocean (19).

The two UCYN-A sub-lineages had similar periodicity of transcript levels to each other, despite divergence in gene sequences at the amino-acid level (average 14% genome-wide), cell morphology (19) and genome size (Figure 1). There were four gene clusters based on the time of day exhibiting the highest relative transcript level (Figure 1). Cluster I had the highest relative transcript level during the day (with a maximum 10 h into the light period) and included genes involved in cell division (e.g. *ftsZ*, *murG*, *minE*, *murB*), DNA replication (e.g. *topA*, *rpoE*, *DPO3B*), ABC transporters (e.g. *nikA*, *nikB*, *pstC*, *cbiO*), carbohydrate and lipid metabolism (e.g. *pdhA*, *pgi*, *fabG*, *fabH*) and a few photosynthesis genes (*petL*, *psaD and ccsB*). The transcripts for

the *petL* gene, encoding subunit 6 of the cytochrome b_6f complex and the only nitrogen fixation-related gene in this cluster (*nifK*) had a substantial change at this time (more than 3-fold).

The transcript abundance of genes from clusters II and III had similar patterns, with an increase before sunrise and a decrease during the dark period. The highest relative transcript levels for clusters II and III were 4h and 1h after sunrise, respectively, and included genes involved in nitrogen fixation (*nifHDK* operon) that increased 4-fold during the light period. However, these clusters also included genes involved in oxidative phosphorylation (e. g., NADH dehydrogenases subunits and ATP synthase related genes), carbohydrate catabolism such as those involved in glycolysis (e.g. *gap1, fbaA, pgi, eno*), the pentose phosphate pathway (*opcA* and *zwf*) and photosynthesis (e.g. cytochrome b_6f complex subunit genes). In most cyanobacteria, genes encoding proteins involved in carbohydrate catabolism are highly transcribed during the night and are essential for survival under dark conditions.

The gene with the most dramatic difference in transcript levels between the light and dark periods encoded the membrane protein COP23 (23 kDa circadian oscillating protein), which had more than a 5-fold change in transcript abundance in both UCYN-A strains (Figure 1). COP23, a protein which may have a critical role in membrane function, has only been detected in nitrogen-fixing cyanobacteria (20).

Cluster IV had genes with the highest transcript level during the night and lowest during the day and included genes encoding photosystem I (PSI) subunits, a carbohydrate porin (*oprB*) and also genes encoding ribosomal proteins with 2- and 4-fold changes during the night period. Cluster IV had the lowest number of genes compared with the other clusters. Surprisingly, the PSI genes (*psaA* and *psaB)* were expressed during the night as in many anoxygenic phototrophic bacteria (21), whereas in most oxygenic cyanobacteria (including mats) these genes are expressed during the day (22).

The results show that UCYN-A has a daily rhythm of gene expression with strong periodicities of transcript levels over the diel cycle. Daily patterns of gene transcription in cyanobacteria are typically regulated by a circadian rhythm mediated by *kai* gene products (11). Rhythmic daily transcription patterns are still possible without the full suite of *kai* genes, for example, the marine cyanobacterium *Prochlorococcus* sp. MED4 lacks one of the circadian genes, *kaiA*, yet it maintains strong diel gene transcription patterns (18). However, *Prochlorococcus* sp. PCC 9511 loses the typical periodicities of the circadian clock under continuous light (23). In the case of UCYN-A, it lacks two of the three *kai* genes (24), which is unique among cyanobacteria, and furthermore, the *kaiC* gene was not transcribed at detectable levels. It is unclear what controls the UCYN-A diel gene expression pattern, but it could be that 1) there are unidentified components of a clock and signal transduction pathway, or that 2) the pattern could be driven by the physiological differences between light and dark conditions, which might be primarily driven by energy supplied by the eukaryotic partner. It is possible that the diel transcription patterns in UCYN-A are primarily regulated by the daily host metabolism, which itself is likely to be circadian. However, it is not yet known whether the UCYN-A diel cycle is maintained under constant conditions in UCYN-A, or whether the diel pattern is maintained in the absence of the partner alga.

UCYN-A transcription patterns are similar to aerobic marine daytime N_2-fixers and non- N_2-fixers

UCYN-A had diel whole genome expression patterns that were different from those of phylogenetically closely related unicellular cyanobacteria (17). Only a few genes (such as those encoding ATP synthase) had the same daily pattern among all cyanobacteria, presumably differing because of physiology (e.g. N_2-fixing or not). The unicellular cyanobacteria *C. watsonii* WH 8501 and *Cyanothece* sp. ATCC 51142, which fix N_2 during the night, expressed many genes in an

187 opposite pattern compared to the day-time N_2-fixing *T. erythraeum* and UCYN-A (Figure 2 and
188 Tables S4 and S5). Interestingly, the diel transcription patterns of N_2 fixation and PSI genes in
189 UCYN-A were opposite to those in *Cyanothece* sp. ATCC 51142 and *C. watsonii* WH 8501 and
190 more similar to those of *T. erythraeum* (Figure 2 and Tables S4 and S5).

191 As observed for the activity of nitrogenase, it has been demonstrated that levels
192 of *nif* transcripts and the biosynthesis of different components of the nitrogenase complex are very
193 sensitive to O_2 (22, 25-27), most likely to avoid energy losses associated with the degradation of
194 this enzyme by O_2. Thus, the different patterns observed in the genes involved in N_2-fixation in the
195 cyanobacteria studied here presumably are due to the different mechanisms used to protect the
196 nitrogenase complex from the O_2 produced by photosynthesis. *T. erythraeum* and UCYN-A had
197 the maximum transcript levels of the nitrogenase and PSI genes just prior to dawn, but maintained
198 high levels of transcripts for both sets of genes during the day. The peak of transcript levels just
199 before dawn is likely due to the advantage of synthesizing nitrogenase in preparation for N_2
200 fixation in the early hours of the day (28).

201 The diel expression patterns of genes that are unrelated to N_2 fixation in the aerobic day-time
202 N_2-fixers (*T. erythraeum* and UCYN-A) were also more similar to those of non-N_2-fixing
203 sympatric cyanobacteria of the genus *Prochlorococcus* and to heterocysts of heterocyst-forming
204 cyanobacteria than to the nighttime N_2-fixing cyanobacteria (*C. watsonii* and *Cyanothece* sp.). The
205 transcript levels of genes encoding ribosomal proteins in both UCYN-A and *T. erythraeum* were
206 higher during the night, probably because the reduced nitrogen required for the synthesis of new
207 proteins was obtained during the day (Figure 2 and Tables S4 and S5). Similar patterns were
208 observed in *Prochlorococcus* with higher transcript levels during the night (Figure 2 and Tables S4
209 and S5) while genes encoding ribosomal proteins in *C. watsonii* WH 8501 and *Cyanothece* sp.
210 ATCC 41142 had maximum transcript levels during the day (Figure 2 and Tables S4 and S5).
211 Intriguingly, these results imply that both UCYN-A and *T. erythraeum* have adopted day-time

gene transcription patterns for the main metabolic pathways minimizing cellular processes in the dark. The night-time patterns of the transcript levels of the ribosomal proteins (genes) would make it possible to have proteins synthesized in order to make the most efficient use of the light period, as in *Prochlorococcus*. Because UCYN-A and *Trichodesmium* are likely to be the two most abundant N_2-fixing cyanobacteria in the open ocean, it appears that direct coupling of N_2 fixation to photosynthesis is important in the oligotrophic environment (as long as low oxygen concentrations are maintained in the cell).

Phosphorus is a vital element for cellular energetics and growth and is acquired by oceanic bacterioplankton primarily as phosphate (29-31). The UCYN-A phosphate ABC transporter had the same diel pattern as in *Trichodesmium* for genes involved in DNA replication, with higher transcript levels during the day (Table S5), but maximum transcript abundances during the late afternoon in *Crocosphaera* and *Cyanothece* (17, 32). High levels of phosphate transporters during the day could meet the increased demand for inorganic phosphate (33, 34) during DNA replication, which occurs during the day in UCYN-A and *Trichosdesmium*. Similar patterns were observed in the heterocyst-forming *Richelia* with peak expression of P acquisition genes at approximately 15:00, suggesting the apparent rhythmicity of P acquisition could be a common feature of daytime N_2-fixers (35).

The initiation factor of DNA replication, DnaA, is a protein highly conserved in prokaryotes although it is absent in red algae, the cyanobacterial symbiont *Nostoc azollae* (36) and also the spheroid bodies of diatoms (37). The genome of UCYN-A lacks the *dnaA* gene as well. Recent studies suggested that DnaA is not essential for DNA replication and the lack of *dnaA* could suggest a preadaptation of the genome to enable the symbiosis (38). In UCYN-A and *T. erythraeum*, genes for DNA replication (*dnaE and* RNaseHI), DNA topoisomerases, DNA gyrases and cell division (*ftsZ*, *mre*, *min*) had maximum transcript levels during the day (i.e., after midday), and minimum levels at night (Figure 3A and Figure S1). In contrast, the nighttime N_2-

237 fixing *Cyanothece* sp. ATCC 51142 and *C. watsonii* WH 8501 confine cell division to the period
238 of transition from dark to light at sunrise. The temporal delay in cell division in *Cyanothece* and
239 *Crocosphaera* has been suggested to reflect the need to recover energy reserves with light-derived
240 energy after night-time metabolic activity (39). The similarity of the pattern in UCYN-A to
241 *Trichodesmium* is consistent with UCYN-A shifting metabolism to the daytime.

242 Microscopy counts of the *B. bigelowii* -UCYN-A2 symbiosis were performed eight times
243 during two diel cycles in order to observe the timing of cell division (Figures 3B and C and Table
244 S6). In both diel cycles, single host cells with two associated UCYN-A2 cells (or groups of cells),
245 corresponding to approximately 60% of total cell counts, were present at night between 21:00 and
246 03:00. The delay observed between the higher transcription levels after midday and actual cell
247 division at 21:00 may be explained by the need of the cell to coordinate the assembly of the cell
248 division machinery prior to cell division.

249

250 **Unique UCYN-A transcription patterns**

251 Although many gene transcription patterns in UCYN-A are more similar to
252 *Trichodesmium* than to other unicellular N_2-fixing cyanobacteria, some of the patterns were
253 unique to UCYN-A. Such unique gene transcription patterns in the UCYN-A symbiosis may
254 provide clues to possible roles of specific genes involved in adaptation to N_2-fixing symbiosis
255 revealing metabolic interdependence between host and symbiont. In order to compare the
256 transcriptomic patterns of these specific genes with the rest of the N_2-fixers, we performed
257 network analysis of these genes using Pearson correlation. Whereas most of the key genes of
258 the major pathways in UCYN-A had higher transcript levels during the day, the other
259 unicellular N_2-fixing cyanobacteria had maximum transcript levels at night (Figure 4). For
260 example, glycolysis genes in UCYN-A had the highest levels of transcripts at sunrise and
261 midday (maximum light conditions) in contrast to the other cyanobacteria (Figures 4 and 5).

The metabolic pathway that generates reductant for biosynthesis activities (NADPH), the pentose phosphate pathway (PPP), had similar patterns. The allosteric effector *opcA*, which redirects carbon flow to the first enzyme of the PPP (glucose-6-P dehydrogenase (*zwf*)) (18, 40), had a different periodic transcript level pattern in UCYN-A (Figures 4 and 6) compared to other cyanobacteria (41, 42).

N_2 fixation in UCYN-A depends on the light period for the supply of photosynthate from the host during the day, as well as possibly producing ATP by cyclic photophosphorylation with PSI. Because UCYN-A cannot fix carbon dioxide, it has to obtain reduced carbon compounds in the same way. Based on genome and transcriptomic profiles, we propose a pathway of carbon metabolism for the regeneration of reductant and ATP in UCYN-A, which is needed for N_2 fixation (Figure 7). Carbohydrate porins or ABC transporters could transport the carbohydrates from the host to the cyanobacteria during the day and the carbon compounds metabolized through the oxidative pentose phosphate (OPP) or glycolysis pathways. Pyruvate is required for generation of reductant for nitrogenase and also to generate acetyl-CoA for synthesis of fatty acids.

Because UCYN-A lacks photosystem II, which normally supplies electrons to photosystem I by splitting water, UCYN-A needs alternative electron donors if it uses PSI to make the reductant NADPH. The NADH generated by the OPP pathway or by glycolysis could reduce the plastoquinone pool via the NDH-1 complex and transfer electrons to ferredoxin though the PQ pool, cytochrome b_6f plastocyanin and the action of PSI. Ferredoxin could deliver electrons to the ferredoxin:NADPH oxidoreductase (FNR), which might provide reductant and ATP directly to the dinitrogenase reductase. To increase the ATP/e^- ratio, UCYN-A can redirect electrons from PSI to NDH-1 in cyclic phosphorylation. This mechanism to supply nitrogenase with electrons was proposed years ago for heterocysts (43).

Together, the results are consistent with the assumption that UCYN-A uses host-supplied carbohydrates during the day while other unicellular cyanobacteria synthesize their own carbohydrates during the day and use them during the evening or at night. The unique distribution of these metabolic processes suggests that UCYN-A has developed the ability for light-driven, day-time N_2 fixation under oxic conditions as a result of symbiosis.

Apart from fixed carbon, several other compounds may be made available to UCYN-A, which may be endosymbiotic, and relies on the host for all of its essential nutrients. Interestingly, UCYN-A has the whole pathway for the synthesis of the cyanobacterial type of vitamin B_{12}, pseudocobalamin, that can be required for the activity of several vital enzymes in central metabolism (44) (Table S8). Transcription of genes involved in B_{12} synthesis were detected in all cyanobacteria, and some of them had diel patterns (Table S2 and S8). It is unknown if UCYN-A has enzymes that require pseudocobalamin or whether it can be used by the host. However, in order for the host to use pseudocobalamin, it would have to be remodeled in order to be accessible to the haptophyte (45). The role of pseudo-B_{12} biosynthesis in UCYN-A is unclear, but the fact that UCYN-A retains this entire pathway, in such a reduced genome, indicates that it is likely to have an important role, perhaps in symbiosis.

It is still unclear how N_2 fixation in UCYN-A avoids the oxygen evolved by the photosynthetic host alga. There are only two possible pathways for consuming O_2 in UCYN-A, including aerobic (cytochrome-dependent) respiration and the photocatalyzed reduction of O_2 to H_2O in PSI which occurs in the heterocysts of cyanobacteria like *Nostoc* sp. PCC 7120 (46-48). The latter, called the Mehler reaction, results in the production of the superoxide radical O_2^-, which is subsequently reduced to water (49, 50).

In UCYN-A, the cytochrome *c* oxidase *coxA* gene was transcribed during the night (cluster IV) but also rarely during the day, along with a few N_2 fixation genes (cluster I) (Figure 1). Moreover, we also found higher transcript levels during the day for the antioxidant enzyme

311 superoxide dismutase (*sod1*) and two peroxiredoxins (*prxR*), which have the ability to detoxify
312 peroxide (Figures 1 and 7). Both antioxidants would protect the nitrogenase against the reactive
313 oxygen species produced by UCYN-A or the haptophyte host (Figure 7).

314 It is not currently possible to directly determine the oxygen protection mechanisms in this
315 uncultured microorganism because 1) transcription cannot necessarily be related to function and 2)
316 it is not possible to do physiological experiments with this low-abundance microorganism that has
317 yet to be obtained in an axenic culture. Consequently, the question of protection from O_2 cannot be
318 directly addressed experimentally, but our results suggest that some of the proteins in UCYN-A
319 could help to protect nitrogenase from the O_2 generated by host photosynthesis.

320 Because UCYN-A has genome reduction normally associated with endosymbiosis (e.g. in
321 *Paulinella chromatophora*; (51)), the unique gene transcription patterns of UCYN-A may provide
322 insights into the evolution of endosymbiosis and organellar evolution. Future studies are needed to
323 determine if the rhythm of these patterns is maintained under constant conditions as in a circadian
324 rhythm, whether the host has a circadian rhythm and/or the daily cycle in UCYN-A simply
325 responds to metabolite availability from the host. It will also be interesting to determine how PSI
326 is involved in supporting the energy or reductant requirements of N_2 fixation. Such experiments
327 will have to await the establishment of a pure culture.

328

329 **Materials and Methods**

330 *Diel sampling of UCYN-A*

331 Surface seawater samples for UCYN-A transcription and catalyzed reported deposition-
332 fluorescence *in situ* hybridization (CARD-FISH) analyses were collected using a bucket from the
333 end of the Scripps Institution of Oceanography (SIO) Ellen Browning Scripps Memorial Pier in La
334 Jolla, CA, USA. Two replicates were collected from the bucket at each time point within 48 hours
335 between 28th July and 1st August 2014 for transcriptomic analysis and between 3rd - 8th May 2016

for CARD-FISH. A total of 16 samples were collected every 3-6 hours (two replicates taken at each of eight time points): 12:00-L6, 15:00-L9, 21:00-D3, 00:00-D6, 06:00-2D12, 09:00-2L3, 15:00-2L9 and 18:00-2L12. L and D stand for light and dark period, respectively, 2L and 2D the second light-dark cycle, and the number the corresponding hours entering light or dark period.

For the CARD-FISH assay, from each seawater replicate, 190 mL of seawater was fixed with 10 mL 37% formaldehyde (1.87% v/v final concentration) at 4ºC in the dark for 1 hour. After fixation, 100 mL was filtered at a maximum vacuum pressure of 100 mm Hg onto a 0.6 µm pore-size, 25 mm diameter polycarbonate membrane filter (Millipore Isopore^{TM}, EMD Millipore, Billerica, MA, USA) with a support filter of 0.8 µm pore-size, 25 mm diameter polycarbonate cellulose acetate membrane filter (Sterlitech Corporation, Kent, WA, USA). The filters were kept at -80ºC until processed.

Samples for RNA extraction were collected by filtering a total of 500 mL from each seawater replicate through 0.22 µm pore-size, 47 mm diameter Supor filters (Pall Corporation, Port Washington, NY, USA) using a peristaltic pump. Filters were placed in sterile 2 mL bead-beating tubes with sterile glass beads, flash-frozen in liquid nitrogen and stored at -80ºC until extraction.

Double CARD-FISH assay

The double CARD-FISH assay was carried out following the protocol designed by Cabello et al. 2016 and Cornejo-Castillo et al. 2016. All of the probes, competitors and helpers used in this work are compiled in Table S7. More details are described in Supplementary Information. Microscopic evaluation and counting was performed with the Carl Zeiss Axioplan-2 Imaging Fluorescent Microscope (Zeiss, Berlin, Germany) in 3 transects (8.0 x 0.1 mm^2 each) across the filter piece. Cell dimensions were estimated using AxioVision 4.8 and Image J software (52).

Diel sampling of Trichodesmium erythraeum IMS101 cultures

Biological triplicate cultures of *T. erythraeum* were grown in rectangular canted neck

polycarbonate cell culture flasks with a 0.2 µm pore-size vent cap and 225 cm^2 surface area (Corning Inc., Corning, NY, USA). The cultures were maintained at 26ºC on a 12h:12h light:dark cycle at 50 µmol quanta m^{-2} s^{-1} in YBCII media (53) supplemented with 2.8 µmol L^{-1} ferric ammonium citrate. The light was set on at 7:00 and off at 19:00 hours. The cultures were 10-fold diluted from the inoculum and were verified to be axenic by staining with DAPI and visualizing cells under an epifluorescence microscope (Carl Zeiss, Thornwood, NY, USA). Growth and cell density were monitored until the cultures reached exponential phase (~10-14 days after inoculation), during which the cells were harvested for the diel transcription assay. Samples were taken at 3 hours intervals starting at the onset of the light period until the end of the dark period for a total of 24 hours. A total of 27 samples were collected from these nine time points: 7:00-D12, 10:00-L3, 13:00-L6, 16:00-L9, 19:00-L12, 22:00-D3, 1:00-D6, 4:00-D9 and 7:00-2D12, where L and D stand for light and dark period, respectively, 2D the second light-dark cycle, and the number the corresponding hours entering light or dark period. At each time point, 200 mL each of triplicate cultures (replicates from different flasks) was filtered onto a 5 µm pore-size, 47 mm diameter polycarbonate membrane filter (Osmonics, Minnetonka, MN, USA). The filters were immediately frozen in liquid nitrogen and stored at -80º C until processing.

RNA extraction and processing for hybridization to the microarray

Environmental RNA containing transcripts from UCYN-A cells was extracted using the Ambion RiboPure Bacteria kit (Ambion®, ThermoFisher), with modifications that included mechanical lysis using glass beads (Biospec, Bartlesville, OK). The extracted RNA was treated with Turbo-DNA-free™ DNase Kit (Ambion®, ThermoFisher) to remove genomic DNA. Sufficient environmental RNA was obtained for two replicates at 4 sampling times (L6, L9, D3 and 2L12): L6-1, L6-2, L9-1, L9-2, D3-1, D3-2, 2L12-1 and 2L12-2. L and D stand for light and

385 dark period, respectively, 2L and 2D the second light-dark cycle, and the number the
386 corresponding hours entering light or dark period.

387 Total RNA for *T. erythraeum* was extracted using the Ambion RiboPure Bacteria kit
388 (Ambion®, ThermoFisher), followed by in solution DNase digestion with the RNase-free DNase
389 kit and on-column cleanup with the RNeasy MiniElute kit (Qiagen, Valencia, CA, USA).

390 RNA purity, concentration and quality were determined using a NanoDrop 1000 (Thermo
391 Scientific, Waltham, MA, USA) and a 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA,
392 USA) using the RNA 6000 Nano kit (Agilent Technologies). Only samples with RNA Integrity
393 Number >7.0 and ratios of A260/A230 and A260/A280 ≥1.8 were processed further.

394 From environmental RNA samples that contained UCYN-A, double-stranded (ds) cDNA
395 was synthesized and amplified following the procedure described in Shilova et al..(54). Briefly,
396 400 ng RNA from each sample was used, and 1 µL of 1:100 dilution (corresponding to 4.7
397 attomoles of ERCC-0016) of the (External RNA Control Consortium, (55)) RNA spike-in mix 1
398 (Ambion®) was added before amplification to monitor the technical performance of the assay (55).

399 Double-stranded cDNA was synthesized and amplified using the TransPlex Whole
400 Transcriptome Amplification kit (WTA-2, Sigma-Aldrich, St Louis, MO, USA) and antibody-
401 inactivated hot-start Taq DNA Polymerase (Sigma-Aldrich). The amplified cDNA was purified
402 with the GenElute PCR cleanup kit (Sigma-Aldrich), and the quality and quantity of ds-cDNA was
403 determined with NanoDrop 1000 and a 2100 Bioanalyzer using the Agilent DNA 7500 kit
404 (Agilent Technologies). Four hundred ng of total RNA yielded on average 12 µg of ds-cDNA. The
405 labeling and hybridization of cDNA samples (1.0 µg of ds-cDNA) to the microarray was done at
406 Roy J. Carver Center for Genomics (CCG) Facility (University of Iowa, Iowa city, Iowa, USA)
407 according to the Agilent Technology for arrays protocol.

408 For *T. erythraeum,* at least 30 µg of unamplified total RNA with a concentration of 1.0 µg
409 μL^{-1} per sample was provided for 27 samples. A control sample was generated by mixing equal

amount of total RNA, based on NanoDrop measured concentration, from each of the 27 samples resulting in 28 samples in total. Reverse transcription of the total RNA, labeling of cDNA, and hybridization to the array were performed at the Roche NimbleGen facility according to the manufacturer's protocol (Roche NimbleGen, Inc., Madison, WI, USA).

Design of the UCYN-A array

The oligonucleotide expression array of UCYN-A was designed using UCYN-A1 and UCYN-A2 genes using eArray web-based tool (Agilent Technology Inc.; https://earray.chem.agilent.com/earray/) similar to the array design described in Shilova et al.(54). The gene sequences were obtained from the National Center of Biotechnology Information (NCBI, http://www.ncbi.nlm.nih.gov). Briefly, six probes of 60 nucleotides (nt) length were designed for each gene, and a total of 6618 probes (1199 genes) and 6862 probes (1246 genes) were designed for UCYN-A1 and UCYN-A2, respectively. These probes were replicated (4 times in the 8x60K array slides and 13 times in the 4x180K array slide) which allowed internal evaluation of signals. The sequences of all oligonucleotide probes were tested *in silico* for possible cross-hybridization as described below. The probe sequences were used as queries in the BLASTN against the available nt databases in June 2014: Marine microbes, Microbial Eukaryote Transcription and Non-redundant Nucleotides in the Community Cyberinfrastructure for Advanced Microbial Ecology Research and Analysis (CAMERA, http://camera.calit2.net/,(56)). Agilent technology allows 5% nt mismatch in the whole probe region, thus sequences with a range of 95–100% nt identity to the target probe are detected. Therefore, all probes with BLASTN hits with ≥95% over 100% nt length were deleted. Next, probe sequences that passed the cross-hybridization filter, were clustered using CD-HIT-EST(57, 58) at 95% nt similarity to select unique probes for UCYN-A1 and unique probes for UCYN-A2. Finally, to select probes specific for each strain, the probes with ≥95% nt identity to the genes in the other strain were deleted. However, a few probes that showed cross-hybridization between both strains for highly conserved genes (such as the

435 nitrogenase gene, *nifH*) were retained. In summary, 6120 probes for 1194 genes of UCYN-A1 and
436 6324 probes for 1244 genes of UCYN-A2 were chosen.

437 In addition, standard control probes as part of the Agilent Technology Array (IS-62976-8-
438 V2_60Kby8_GX_EQC_201000210 with ERCC control probes added) were included randomly to
439 feature locations on the microarray slide. The final design of the microarray was synthesized on
440 two platforms: ca. 62976 experimental and 1319 control probes on the 8x60K array slide and ca.
441 180880 experimental and 4854 control probes on the 4x180K array slide. The probe sequences are
442 available at NCBI Gene Expression Omnibus (GEO) under accession number GSE100124.
443
444 *Design of the T. erythraeum IMS101 array*
445 A custom oligonucleotide array for *T. erythraeum* was designed using the Roche NimbleGen
446 platform: (NimbleGen design ID: 080610_Trich_erth_UCSC_TS_expr) according to the complete
447 genome assembly of *T. erythraeum* IMS101 (NC_008312). The genome sequence is publically
448 available via gateways including GenBank (https://www.ncbi.nlm.nih.gov/nuccore/NC_0083120),
449 IMG
450 (http://img.jgi.doe.gov:80/cgibin/pub/main.cgi?section=TaxonDetail&page=taxonDetail&taxon_oi
451 d=637000329), and UCSC genome browser (http://microbes.ucsc.edu/cgi-
452 bin/hgGateway?db=tricEryt_IMS101). Up to six 60-nt long tiling probes were designed to
453 target each of the 4788 genes in the genome, resulting in a total of 28235 probes. The probes were
454 duplicated on the array to allow internal evaluation of hybridization signals. Moreover, tiling 60 nt
455 oligonucleotide probes were also designed to target the intergenic regions >60 bp in length at a
456 150 bp interval, leading to a total of 11175 probes targeting 3877 intergenic regions (average 2.9
457 probes per intergenic region), however hybridization data for intergenic probes are not presented
458 here. All the probes were rank ordered and selected based on the following criteria: 1) they must
459 have a minimum annealing temperature of 68ºC; 2) there is no cross contamination among the

460 probes for different genes and for different intergenic regions. In addition to the experimental
461 probes, standard control probes were also included on the microarray for quality assessment of the
462 sample preparation, the hybridization process and the intensity measurements. The final
463 microarray slides were printed in 4-plex (4x72K) format with 67645 experimental probe features
464 and 7454 control probe features on one array. The full microarray platform descriptions and data
465 for *T. erythraeum* are available at NCBI GEO under accession number GSE99896. Microarray
466 hybridization signals were quantified using a GenePix 4000B Scanner (Molecular Devices,
467 Sunnyvale, CA, USA) at the Roche NimbleGen facility.
468
469 *Microarray data analysis*
470 All data analyses were performed with R (www. R-project.org) and the Bioconductor
471 Project(59), specifically using the Biobase(60), Linear Models for Microarray LIMMA (61),
472 arrayQualityMetrics(62), affyPLM(63, 64), and genefilter packages.
473 *1) UCYN-A microarray*
474 Transcription values for each gene were obtained using median polish summarization, and
475 values were normalized using quantile normalization (63, 64) (Figure S2). The transcription values
476 for UCYN-A at L6, L9, D3 and 2L12 are the mean transcription of the two replicates (L6-1, L6-2,
477 L9-1, L9-2, D3-1, D3-2, 2L12-1 and 2L12-2). Raw and normalized microarray data for UCYN-A
478 were submitted to NCBI GEO under accession number GSE100124. To determine if transcription
479 of a gene was detected, the signal-to-noise ratio (SNR) of each chip was calculated as: SNR = (S_i–
480 BG)/BG; where S_i is the hybridization signal for the gene and BG is the chip background signal
481 determined as average of the lowest 5% of all signals. Transcription was considered detected if
482 SNR of a transcript was ≥5 (as in (Shilova et al. 2014) Transcription values were centered and
483 scaled across genes and samples, and a distance matrix was calculated using Pearson's correlation
484 coefficient. The distance matrix was then used in hierarchical clustering by a complete

485 agglomeration method to identify clusters of genes with similar patterns of transcription during the
486 diel transcription.

487 *2) T. erythraeum microarray*

488 The raw microarray data for *T. erythraeum* were subjected to robust multichip average
489 (RMA) analysis (65) and quantile normalization (63, 64) (Figure 3S). Transcription values for
490 each gene were obtained using median polish summarization (54). Final transcription value for
491 each sample was a mean of up to twelve technical replicates (Blocks 1 and 2 with up to six
492 replicate probes in each block in the *T. erythraeum* microarray design). A gene was selected for
493 further analysis if it had $\log_2$ transcription above 64 in at least 25% of samples and an interquartile
494 range across all samples on the $\log_2$ scale of at least 0.5. This filtering resulted in 4128 genes,
495 which were used in further analysis.

496 *3) Comparison of diel transcription patterns for all cyanobacteria*

497 Transcription data for *Prochlorococcus* sp. MED4, *Cyanothece* sp. ATCC 51142 and
498 *Crocosphaera watsonii* WH 8501 was collected from previous published data (16-18).
499 *Cyanothece* sp. ATCC 51142 and *C. watsonii* WH 8501 microarray data were downloaded
500 from ArrayExpress (http://www.ebi.ac.uk/aerep/) using accession no. E-TABM-386 and E-
501 TABM-737, respectively. The genes with periodic transcriptional patterns for all studied
502 cyanobacteria (*Prochlorococcus* sp. MED4, *Cyanothece* sp. ATCC 51142, *C. watsonii* WH
503 8501, *T. erythraeum* and UCYN-A) were identified using the R package "cycle" based on
504 Fourier analysis, and the genes with FDR<0.25 were selected for further comparison (66)
505 (Table S2). To compare the diel transcription patterns among the cyanobacteria, gene
506 transcription values for each cyanobacterium were selected for over 36 hours. Eight points
507 were selected for UCYN-A (L6, L9, D3, D6, 2D12, 2L3, 2L9, 2L12), 9 points for *T.*
508 *erythraeum* (D12, L3, L6, L9, L12, D3, D6, D9, 2D12), 6 points for *Cyanothece* sp. ATCC
509 51142 (L2, L6, L10, D2, D6, D10), 8 points for *C. watsonii* WH 8501 (D11, L1, L6, L11, D1,

510 D6, 2D11, 2L1) and 19 points for *Prochlorococcus* sp. MED4 (D12 - 2L12 every 2 hours). L
511 and D stand for light and dark period, respectively, 2L and 2D the second light-dark cycle, and
512 the number the corresponding hours entering light or dark period. Because the studies had a
513 few dissimilar sampling times, the missing values were interpolated using the Stineman
514 algorithm implemented in the *imputeTS* package (67). A network was constructed based on the
515 Pearson correlation and using 'make_network' function in phyloseq (68). The maximum
516 distance between connecting nodes was selected as 0.5 unless otherwise noted in figure
517 legends.
518

519 **Acknowledgments**

520 We thank J.C. Meeks (University of California, Davis) for discussions, F. Azam (Scripps
521 Institution of Oceanography, UC San Diego) for access to Scripps facilities, and K. Turk-Kubo
522 and M. Hogan for lab and field support. J.M. García-Fernández and J. Díez-Dapena (University of
523 Córdoba, Spain) and Marine Landa (University of Santa Cruz, CA, USA) for helping us to
524 improve the manuscript. Microarray data have been deposited at NCBI Gene Expression Omnibus
525 (GEO) under accession numbers GSE100124 and GSE99896. The following secure token has
526 been created to allow review of record GSE100124 while it remains in private status:
527 ufchawekhzmlpaj.
528

529 **Author contributions**

530 M.M.M. designed the UCYN-A array, designed and performed the research and analyzed the data.
531 I.N.S. analyzed the *T. erythraeum* array data, aided with the design of the UCYN-A array and
532 comparison of transcription among cyanobacteria. T.S. designed the *T. erythraeum* array and
533 performed the diel sampling of *T. erythraeum* cultures. H.F. aided sampling diel UCYN-A
534 samples and performed the phylogenetic tree. A.M.C carried out and counted the CARD-FISH diel

535 samples. J.P.Z. conceptualized the study, and M.M.M., I.N.S., T.S., H.F. and J.P.Z. drafted and
536 edited the manuscript and figures. All authors read and approved the final manuscript.

537

538 **Competing financial interests**

539 The authors declare no competing financial interest.

540

541

542 **References**

543 1. Postgate JR. 1982. The fundamentals of nitrogen fixation. Cambridge University Press,
544 London.

545 2. Montoya JP, Holl CM, Zehr JP, Hansen A, Villareal TA, Capone DG. 2004. High rates
546 of N_2 fixation by unicellular diazotrophs in the oligotrophic Pacific Ocean. Nature
547 430:1027-1031.

548 3. Cabello AM, Cornejo-Castillo FM, Raho N, Blasco D, Vidal M, Audic S, de Vargas C,
549 Latasa M, Acinas SG, Massana R. 2015. Global distribution and vertical patterns of a
550 prymnesiophyte-cyanobacteria obligate symbiosis. ISME J 10:693-706.

551 4. Martínez-Pérez C, Mohr W, Löscher C, Dekaezemacker J, Littmann S, Yilmaz P,
552 Lehnen N, Fuchs BM, Lavik G, Schmitz R, LaRoche J, Kuypers MM. 2016. Small
553 unicellular diazotrophic symbiont is a key player in the marine nitrogen cycle. Nature
554 Microbiology doi:DOI: 10.1038/NMICROBIOL.2016.163.

555 5. Farnelid H, Turk-Kubo K, Muñoz-Marin M, Zehr J. 2016. New insights into the
556 ecology of the globally significant uncultured nitrogen-fixing symbiont UCYN-A.
557 Aquatic Microbial Ecology 77:125-138.

558 6. Tripp HJ, Bench SR, Turk KA, Foster R, Desany BA, Niazi F, Affourtit JP, Zehr JP.
559 2010. Metabolic streamlining in an open-ocean nitrogen-fixing cyanobacterium. Nature
560 464.

561 7. Thompson AW, Foster RA, Krupke A, Carter BJ, Musat N, Vaulot D, Kuypers MMM,
562 Zehr JP. 2012. Unicellular cyanobacterium symbiotic with a single-celled eukaryotic
563 alga. Science 337:1546-1550.

564 8. Church M, Short C, Jenkins B, Karl D, Zehr J. 2005. Temporal patterns of nitrogenase
565 gene (*nifH*) expression in the oligotrophic North Pacific Ocean. Applied and
566 Environmental Microbiology.

567 9. Krupke A, Musat N, LaRoche J, Mohr W, Fuchs BM, Amann RI, Kuypers MMM,
568 Foster RA. 2013. In situ identification and N_2 and C fixation rates of uncultivated
569 cyanobacteria populations. Systematic and Applied Microbiology 36:259-271.

570 10. Thompson A, Carter BJ, Turk-Kubo K, Malfatti F, Azam F, Zehr JP. 2014. Genetic
571 diversity of the unicellular nitrogen-fixing cyanobacteria UCYN-A and its
572 prymnesiophyte host. Environ Microbiol 16:3238-49.

573 11. Cohen SE, Golden S. 2015. Circadian rhythms in cyanobacteria. Microb Mol Biol Rev
574 79:373-385.

575 12. Bell-Pedersen D, Cassone VM, Earnest DJ, Golden SS, Hardin PE, Thomas TL, Zoran
576 MJ. 2005. Circadian rhythms from multiple oscillators: Lessons from diverse
577 organisms. Nature Reviews Genetics 6:544-556.

578 13. Reimers AM, Knoop H, Bockmayr A, Steuer R. 2017. Cellular trade-offs and optimal
579 resource allocation during cyanobacterial diurnal growth. Proc Natl Acad Sci USA
580 doi:https://doi.org/10.1073/pnas.1617508114.

581 14. Golden SS, Ishiura M, Johnson CH, Kondo T. 1997. Cyanobacterial circadian rhythms.
582 Annual Review of Plant Physiology and Plant Molecular Biology 48:327-354.

583 15. Kondo T, Ishiura M. 2000. The circadian clock of cyanobacteria. BioEssays 22:10-15.

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

587 17. Shi T, Ilikchyan IN, Rabouille S, Zehr JP. 2010. Genome-wide analysis of diel gene
588 expression in the unicellular N_2-fixing cyanobacterium *Crocosphaera watsonii* WH
589 8501. ISME J 4:621-632.

590 18. Zinser ER, Lindell D, Johnson ZI, Futschik ME, Steglich C, Coleman ML, Wright MA,
591 Rector T, Steen R, McNulty NP, Thompson LR, Chisholm SW. 2009. Choreography of
592 the transcriptome, photophysiology, and cell cycle of a minimal photoautotroph,
593 *Prochlorococcus*. PLoS ONE 4:e5135.

594 19. Cornejo-Castillo FM, Cabello AM, Salazar G, Sanchez-Baracaldo P, Lima-Mendez G,
595 Hingamp P, Alberti A, Sunagawa S, Bork P, de Vargas C, Raes J, Bowler C, Wincker
596 P, Zehr JP, Gasol JM, Massana R, Acinas SG. 2016. Cyanobacterial symbionts
597 diverged in the late Cretaceous towards lineage-specific nitrogen fixation factories in
598 single-celled phytoplankton. Nat Commun 7:11071.

599 20. Rong-Fong L, Kun-Da T, Tan-Chi H. 2003. Factors affecting the circadian degradation
600 of COP23 in *Synechococcus* RF-1. Botanical Bulletin-Academia Sinica Taipei 44:151-
601 158.

602 21. Klatt CG, Liu Z, Ludwig M, Kühl M, Jensen SI, Bryant DA, Ward DM. 2013.
603 Temporal metatranscriptomic patterning in phototrophic *Chloroflexi* inhabiting a
604 microbial mat in a geothermal spring. ISME Journal 7:1775-1789.

605 22. Steunou AS, Jensen SI, Brecht E, Becraft ED, Bateson MM, Kilian O, Bhaya D, Ward
606 DM, Peters JW, Grossman AR, Kühl M. 2008. Regulation of *nif* gene expression and
607 the energetics of N_2 fixation over the diel cycle in a hot spring microbial mat. The
608 ISME Journal 2:364-378.

609 23. Holtzendorff J, Partensky F, Mella D, Lennon J-F, Garczarek L. 2008. Genome
610 streamlining results in loss of robustness of the circadian clock in the marine
611 cyanobacterium *Prochlorococcus marinus* PCC 9511. J. Biol Rhythms 23:187-199.

612 24. Axmann I, Hertel S, Wiegard A, Dörrich A, Wilde A. 2014. Diversity of KaiC-based
613 timing systems in marine Cyanobacteria. Marine Genomics 14:3-16.

614 25. Fay P. 1992. Oxygen relations of nitrogen fixation in cyanobacteria. Microbiology and
615 Molecular Biology Reviews 56:340-373.

616 26. Staal M, Rabouille S, Stal LJ. 2007. On the role of oxygen for nitrogen fixation in the
617 marine cyanobacterium *Trichodesmium* sp. Envionmental Microbiology 9:727-736.

618 27. Bürgmann H, Widmer F, Sigler WV, Zeyer J. 2003. mRNA extraction and reverse
619 transcription-PCR protocol for detection of *nif*H gene expression by *Azotobacter*
620 *vinelandii* in soil. Appl. Environ. Microbiol. 69:1928-1935.

621 28. Chen YB, Dominic B, Mellon MT, Zehr JP. 1998. Circadian rhythm of nitrogenase
622 gene expression in the diazotrophic filamentous nonheterocystous cyanobacterium
623 *Trichodesmium* sp. strain IMS 101. J Bacteriol 180:3598-3605.

624 29. Zubkov MV, Martin AP, Hartmann M, Grob C, Scanlan D. 2015. Dominant oceanic
625 bacteria secure phosphate using a large extracellular buffer. Nature Commun 6.

626 30. Zubkov MV, Mary I, Woodward EMS, Warwick PE, Fuchs BM, Scanlan DJ, Burkill
627 PH. 2007. Microbial control of phosphate in the nutrient-depleted North Atlantic
628 subtropical gyre. Environ. Microbiol 9:2079-2089.

629 31. Michelou VK, Lomas MW, Kirchman DL. 2011. Phosphate and adenosine-5´-
630 triphosphate uptake by cyanobacteria and heterotrophic bacteria in the Sargasso Sea.
631 Limnol Oceanogr 56:323-332.

632 32. Stockel J, Welsh EA, Liberton M, Kunnvakkam R, Aurora R, Pakrasi HB. 2008. Global
633 transcriptomic analysis of *Cyanothece* 51142 shows a robust diurnal oscillation of
634 central metabolic processes. *Proc Natl Acad Sci* USA 105:6156-6161.

635 33. Davey M, Tarran G, Mills M, Ridame C, Geider R, Roche JL. 2008. Nutrient limitation
636 of picophytoplankton photosynthesis and growth in the tropical North Atlantic. Limnol
637 Oceanogr 53:1722-1733.

638 34. Klausmeier C, Litchman E, Levin S. 2004. Phytoplankton growth and stoichiometry
639 under multiple nutrient limitation. Limnol Oceanogr 49:1463-1470.

640 35. Harke MJ, Frischkorn KR, Haley ST, Aylward FO, Zehr J, Dyhrman ST. 2018.
641 Periodic and coordinated gene expression between a diazotroph and its diatom host.
642 ISME Journal doi:https://doi.org/10.1038/s41396-018-0262-2.

643 36. Ran L, Larsson J, Vigil-Stenman T, Nylander J, Ininbergs K, Zheng W, Lapidus A,
644 Lowry S, Haselkorn R, Bergman B. 2010. Genome erosion in a nitrogen-fixing
645 vertically transmitted endosymbiotic multicellular cyanobacterium. PLoS ONE
646 5:e11486.

37. Nakayama T, Kamikawa R, Tanifuji G, Kashiyama Y, Ohkouchi N, Archibald JM, Inagaki Y. 2014. Complete genome of a nonphotosynthetic cyanobacterium in a diatom reveals recent adaptations to an intracellular lifestyle. *Proc Natl Acad Sci* USA:201405222.

38. Ohbayashi R, Watanabe H, Ehira S, Kanesaki Y, Chibazakura T, Yoshikawa H. 2016. Diversification of DnaA dependency for DNA replication in cyanobacterial evolution. The ISME J 10:1113-1121.

39. Wilson ST, Aylward FO, Ribalet F, Barone B, Casey JR, Connell PE, Eppley JM, Ferrón S, Fitzsimons JN, Hayes CT, Romano AE, Turk Kubo K, Vislova A, Armbrust EV, Caron DA, Church M, Zehr JP, Karl D, DeLong E. 2017. Coordinated regulation of growth, activity and transcription in natural populations of the unicellular nitrogen-fixing cyanobacterium *Crocosphaera*. Nature Microbiol 2.

40. Hagen KD, Meeks JC. 2001. The unique cyanobacterial protein OpcA is an allosteric effector of glucose-6-phosphate dehydrogenase in *Nostoc punctiforme* ATCC 29133. Journal of Biological Chemistry 276:11477-11486.

41. Min HT, Golden SS. 2000. A new circadian class 2 gene, *opcA*, whose product is important for reductant production at night in *Synechococcus elongatus* PCC 7942. J Bacteriol 182:6214-6221.

42. Lee S, Ryu J, SY, Jeon J, Song J, Cho H, Choi S, Choi D, Tandeau de Marsac N, Park Y. 2007. Transcriptional regulation of the respiratory genes in the cyanobacterium *Synechocystis* sp. PCC 6803 during the early response to glucose feeding. Plant Physiol 145:1018-1030.

43. Maldener I, Muro-Pastor AM. 2010. Cyanobacterial heterocysts. *In* Sons JW (ed), Encyclopedia of Life Sciences (ELS) doi:DOI: 10.1002/9780470015902.a0000306.pub2, Chichester, UK.

44. Martens J, Barg H, Warren M, Jahn D. 2002. Microbial production of vitamin B12. Appl Microbiol Biotechnol 58:275-285.

45. Helliwell KE, Lawrence AD, Holzer A, Kudahl UJ, Sasso S, Kräutler B, Scanlan DJ, Warren MJ, Smith AG. 2016. Cyanobacteria and eukaryotic algae use different chemical variants of vitamin B12. Curr Biol 26:999-1008.

46. Milligan AJ, Berman-Frank I, Gerchman Y, Dismukes GC, Falkowski PG. 2007. Light-dependent oxygen consumption in nitrogen-fixing cyanobacteria plays a key role in nitrogenase protection. J Phycol 43:845-852.

47. Valladares A, Herrero A, Pils D, Schmetterer G, Flores E. 2003. Cytochrome c oxidase genes required for nitrogenase activity and diazotrophic growth in *Anabaena* sp. PCC 7120. Mol Microbiol 47:1239-1249.

48. Jones KM, Haselkorn R. 2002. Newly identified cytochrome c oxidase operon in the nitrogen-fixing cyanobacterium *Anabaena* sp. strain PCC 7120 specially induced in heterocyts. J Bacteriol 184:2491-2499.

686 49. Lafiti A, Ruiz M, Cheng-Cai Z. 2009. Oxidative stress in cyanobacteria. FEMS
687 Microbiol Rev 33:258-278.

688 50. Milligan AJ, Berman-Frank I, Gerchman Y, Dismukes GC, Falkowski PG. 2007. Light-
689 dependent oxygen consumption in nitrogen-fixing cyanobacteria plays a key role in
690 nitrogenase protection. J Phycol 43:845-852.

691 51. Nakayama T, Archibald JM. 2012. Evolving a photosynthetic organelle. BMC Mol
692 Biol 10.

693 52. Collins TJ. 2007. Introduction to ImageJ for light microscopy. Microsc Microanal
694 13:1674-1675.

695 53. Chen Y, Zehr JP, Mellon MT. 1996. Growth and nitrogen fixation of the diazotrophic
696 filamentous nonheterocystous cyanobacterium *Trichodesmium* sp. IMS 101 in defined
697 media: evidence for a circadian rhythm J Phycol 32:916-923.

698 54. Shilova IN, Robidart JC, Tripp HJ, Turk-Kubo K, Wawrik B, Post AF, Thompson AW,
699 Ward B, Hollibaugh JT, Millard A. 2014. A microarray for assessing transcription from
700 pelagic marine microbial taxa. The ISME J 8:1476-1491.

701 55. Lemire A, Lea K, Batten D, Gu SJ, Whitley P, Bramlett K. 2011. Development of
702 ERCC RNA spike-in control mixes. J Biomol Tech 22:S46.

703 56. Sun S, Chen J, Li W, Altinatas I, Lin A, Peltier S, Stocks K, Allen EE, Ellisman M,
704 Grethe J, Wooley J. 2011. Community cyberinfrastructure for Advanced Microbial
705 Ecology Research and Analysis: the CAMERA resource Nucl Acids Res 39:D546-
706 D551.

707 57. Huang Y, Niu B, Gao Y, Fu L, Li W. 2010. CD-HIT Suite: a web server for clustering
708 and comparing biological sequences. Bioinformatics 26:680-682.

709 58. Li W, Godzik A. 2006. Cd-hit: a fast program for clustering and comparing large sets
710 of protein or nucleotide sequences. Bioinformatics 22:1658 - 1659.

711 59. Gentleman R, Carey V, Bates D, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L,
712 Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C,
713 Maechler M, Rossini A, Sawitzki G, Smith C, Smyth G, Tierney L, Yang J, Zhang J.
714 2004. Bioconductor: open software development for computational biology and
715 bioinformatics. Genome Biology 5:R80.

716 60. Huber W, Carey JV, Gentleman R, Anders S, Carlson M, Carvalho SB, Bravo CH,
717 Davis S, Gatto L, Girke T, Gottardo R, Hahne F, Hansen DK, Izarry AR, Lawrence M,
718 Love IM, Waldron L, Morgan. 2015. Orchestrating high-throughput genomic analysis
719 with Bioconductor. Nat Meth 12:115121.

720 61. Smyth GK. 2005. Limma: Linear models for microarray data., p 397-420. *In* Gentleman
721 R, Carey V, Dudoit S, Irizarry R, Huber W (ed), Bioinformatics and Computational
722 Biology Solutions using R and Bioconductor. Springer, New York.

723 62. Kauffmann A, Gentleman R, Huber W. 2009. arrayQualityMetrics-a bioconductor
724 package for quality assessment of microarray data. Bioinformatics 25:415-416.

725 63. Bolstad BM, Collin F, Brettschneider J, Simpson K, Cope L, Irizarry RA, Speed TP. 2005. Quality Assessment of Affymetrix GeneChip Data, p 33-47. *In* Springer NY (ed), Bioinformatics and Computational Biology Solutions using R and Bioconductor.

728 64. Bolstad B. 2004. Low level analysis of high-density oligonucleotide Array Data: Background, Normalization and SummarizationUniversity of California, Berkeley.

730 65. Irizarry R, Hobbs B, Collin F, Beazer-Barclay Y, Antonellis K, Scherf U, Speed T. 2003. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 4:249.

733 66. Futschik ME, Herzel H. 2008. Are we overestimating the number of cell-cycling genes? The impact of background models on time series analysis. Bioinformatics 24:1063-1069.

736 67. Moritz S, Sardá A, Bartz-Beielstein T, Zaefferer M, Stork J. 2015. Comparison of different methods for univariate time series imputation in R. arXiv:151003924

738 68. McMurdie P, Holmes S. 2013. Phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. PLoS ONE 8:e61217.
740

741

Titles and legends to main figures

Figure 1. Four different clades based on Pearson correlation of the transcription profile of UCYN-A1 and UCYN-A2 genes over light–dark cycles. The transcription value of each gene at each time point was normalized to the mean at all time points and divided by standard deviation (SD) (*Y* axis, log 2 scale). The *X* axis represents time points where D and L stand for dark and light, respectively, followed by the corresponding hour into the light or dark periods. The second light-dark cycle is shown as 2D followed by the number of the corresponding hours entering light or dark period. The shaded area represents the dark period. In each cluster, most representative genes are listed in the table attached to the plot. UCYN-A1 genes are coded in green and UCYN-A2 genes are coded in orange.

Figure. 2. A. Transcriptional network based on Pearson correlation of gene transcription over the diel cycle in all studied cyanobacteria. The genes are connected if correlation coefficient for their transcription patterns is higher than 0.5. The genes shown are diel genes with variable transcription patterns among the studied cyanobacteria. The arrows point to genes for glycolysis, PPP and N_2 fixation in the studied diazotrophs. The purple circles demarcate genes for ribosomal proteins included in the analysis. Abbreviations: *Prochlorococcus* sp. MED4 (Pro), *Cyanothece* sp. ATCC 51142 (Cya), *C. watsonii* WH 8501 (Cro), *T. erythraeum* (Tricho), Pentose Phosphate Pathway (PPP), Photosystem I (PSI). **B.** Four time course plots are attached for the N_2-fixing cyanobacteria showing the diel transcription patterns of photosystem I, N_2 fixation and genes for ribosomal proteins.

Figure 3. Transcription of genes for replication and cell division in UCYN-A. **Upper Panel: A)** Diel transcription patterns for cell division and replication genes in UCYN-A1 and UCYN-A2 over the light-dark cycle. Hierarchical clustering of genes was based on Pearson correlation

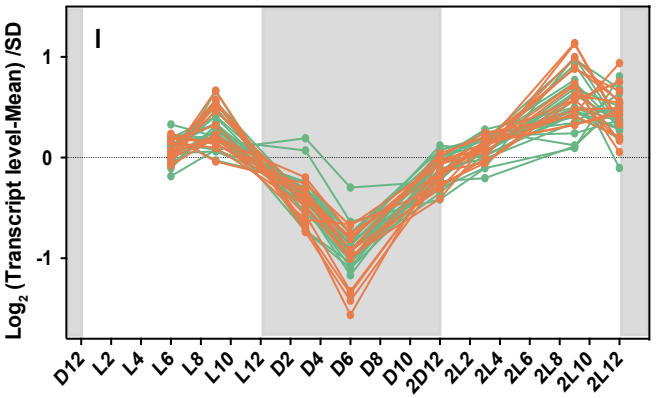
768 between their transcription profiles. The transcription values of genes at each time point were
769 standardized, and the blue-red scale shows how many standard deviations a transcription value
770 was lower or higher, respectively, from the mean transcription values over the diel cycle (Z score).
771 Gene ID and gene product corresponding to each gene for UCYN-A1 and UCYN-A2 are shown.
772 Time is shown on X-axis as light (L) and dark (D), respectively, followed by the corresponding
773 hour after the sunrise and sunset periods started. The second light-dark cycle is shown as 2D
774 followed by the number of the corresponding hours entering light or dark period.
775 **Lower Panel:** Epifluorescence micrographs of dividing UCYN-A2 detected with CARD-FISH
776 (19). **B)** Two big clusters of UCYN-A2 cells and the haptophyte host attached. Left Panel: the
777 nucleus of the host and the UCYN-A2 cells were visualized with-DAPI stain (blue). Right Panel:
778 The UCYN-A2 (red) and its haptophyte host (green). **C)** Two different associations of UCYN-A2
779 with its haptophyte dividing in samples from Scripps Pier.
780
781 **Figure 4.** Network showing the Pearson correlation for gene transcriptions in the unicellular N_2-
782 fixing cyanobacteria *Cyanothece* sp. ATCC 51142 (*Cyanothece*), *C. watsonii* WH 8501
783 (*Crocosphaera*) and UCYN-A. Shown here are key genes in major metabolic pathways with
784 distinct diel transcription patterns. The genes are connected if their correlation coefficient for
785 transcription patterns is higher than 0.2. PPP, pentose phosphate pathway; PSI, photosystem I.
786
787 **Figure 5.** Transcriptional profiles of the genes for glycolysis over light–dark cycles in the
788 cyanobacteria studied here. The transcription value of each gene at each time point was
789 normalized to the mean at all time points and divided by standard deviation (SD) (*Y* axis, log
790 scale). The *X* axis represents time points where D and L stand for dark and light, respectively,
791 followed by the corresponding hour into the light or dark periods. The second light-dark cycle is

792 shown as 2D followed by the number of the corresponding hours entering light or dark period. The
793 shaded area represents the dark period.
794
795 **Figure 6.** Transcriptional profiles of *opcA* (allosteric effector) and *zwf* (glucose-6-P
796 dehydrogenase) over light–dark cycles in the cyanobacteria studied here. The transcription value
797 of each gene at each time point was normalized to the mean at all time points and divided by
798 standard deviation (SD) (*Y* axis, log scale). The *X* axis represents time points where D and L stand
799 for dark and light, respectively, followed by the corresponding hour into the light or dark periods.
800 The second light-dark cycle is shown as 2D followed by the number of the corresponding hours
801 entering light or dark period. The shaded area represents the dark period.
802
803 **Figure 7.** Schematic model of UCYN-A showing the possible main cellular functions, metabolic
804 pathways and transporters.
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819

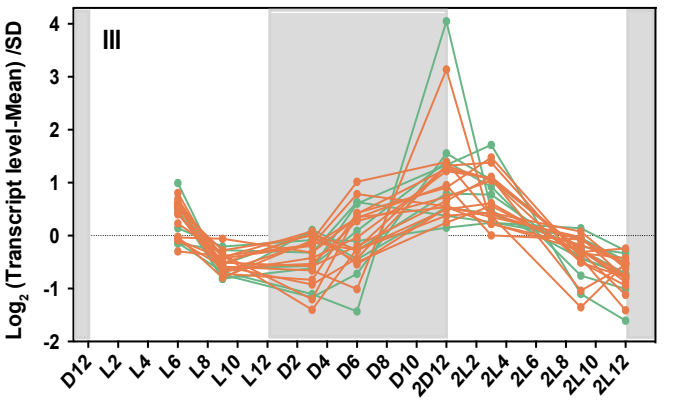
Organism

UCYN-A1
UCYN-A2

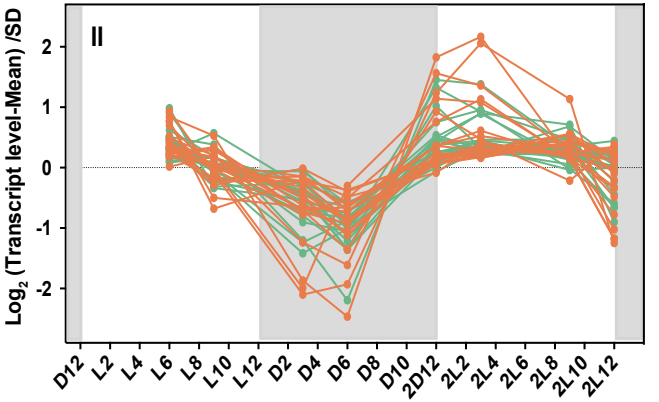
UCYN-A1		UCYN-A2	
fabD	ftsZ	-	-
-	tpi	fabH	-
lpxC	-	lpxC	rpe
sqd1	pdha	-	pdha
hypD	-		thiL
cbio	acp		acp
pstA	glnA		glnA
pstC	g6pd	pstC	-
minD	-	-	accD
-	accA	minE	
-	-	mreC	topA
g6pi	-		rpoE
murG	dpo3	-	-
murE	-	-	sufB
-	fepB	murJ	-
-	fepGD	murB	-



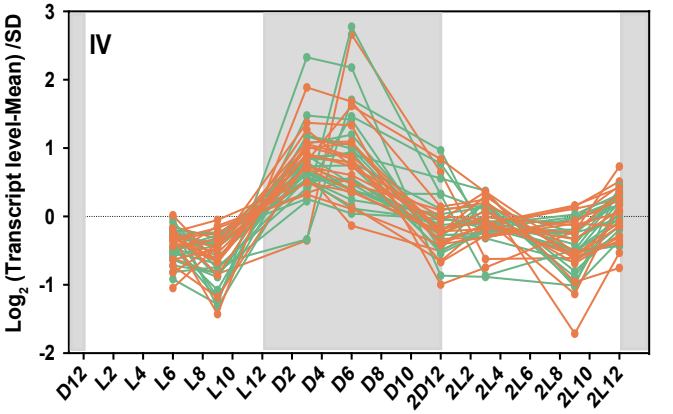
UCYN-A1	UCYN-A2
atpF	atpF
atpB	atpB
-	atpA
pgd	-
rpiA	pstA
cop23	cop23
iscA	iscA
-	eno
-	ndhK
-	ndhC
-	psaJ
-	psaL
-	tktA
-	pgd



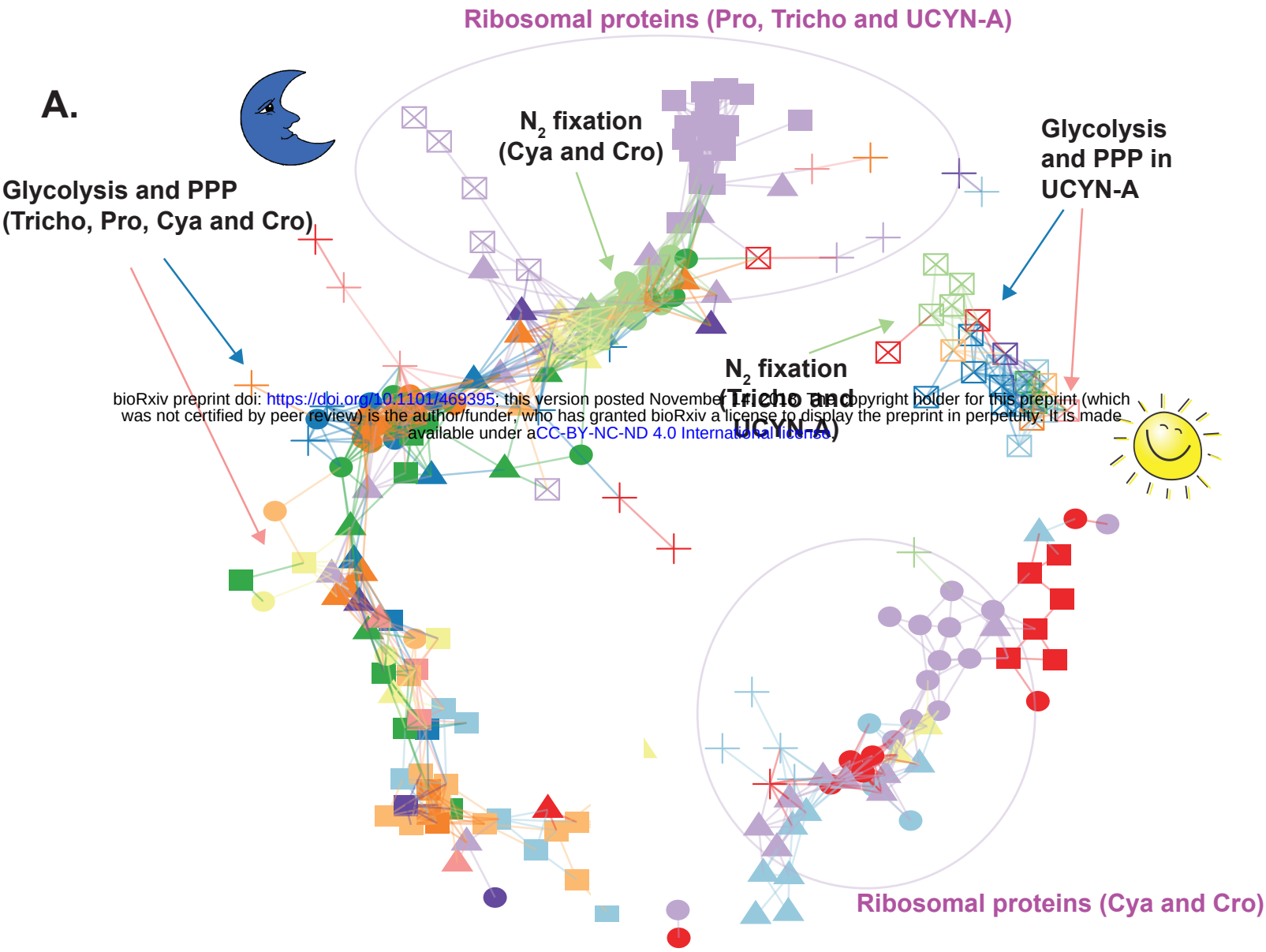
UCYN-A1		UCYN-A2	
ndhJ	petC	ndhJ	petC
ndhD	petF	ndhD	petF
hupS	nikB	-	-
-	sufB	hypE	-
-	-	hypD	petH
g3pd	-	-	lpxA
-	-	fba	lpxB
atpA	fabH	atpA	-
atpI	fabF	atpI	-
-	prpS	atpF	-
atpC	accC	-	-
psaC	-	psaC	gpi
minE	nifU	-	-
-	nifH	ftsZ	nifH
-	nifD	murD	nifD
prxR	sod1	prxR	-
petA	-	petA	nifK



UCYN-A1		UCYN-A2	
psaA	-	psaA	rps13
psaB	-	-	rpl13
oprB	-	oprB	rpl18
rpl11	-	rpl11	rpl5
nusG	-	-	rps9
secE	-	-	rpl24
rps19	-	rps19	rpl2
rpl22	-	rpl22	rpl3
rpl16	-	-	rpl23
rpl24	-	-	rpl29
rps8	-	-	rpl1
rpl6	-	-	guaB
rpl15	-	rpl15	sqdB
dnaE	-	-	acyl
truA	-	-	rpoA
-	-	coxB	-



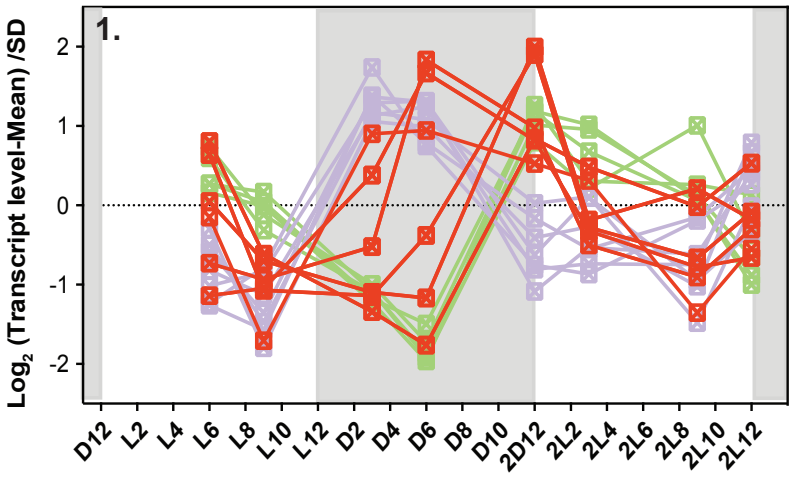
A.

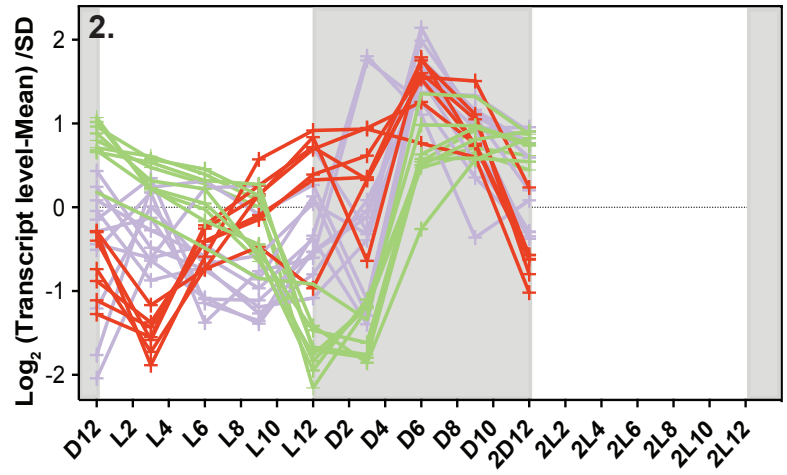


B.

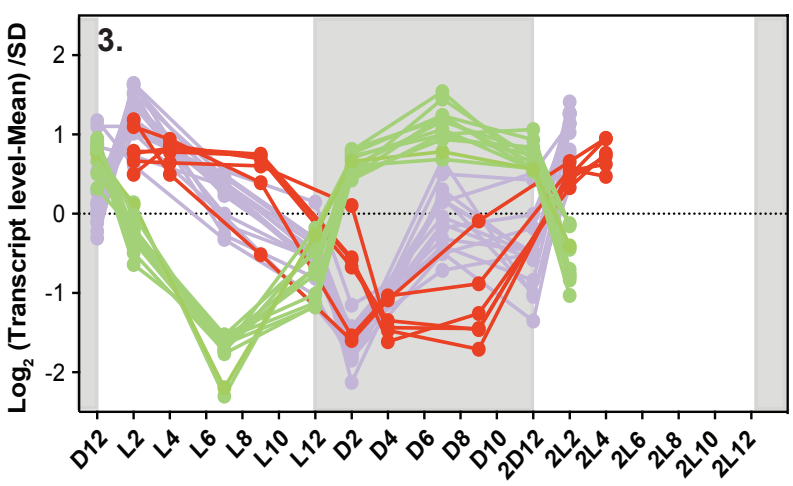
UCYN-A

Trichodesmium erythraeum IMS101

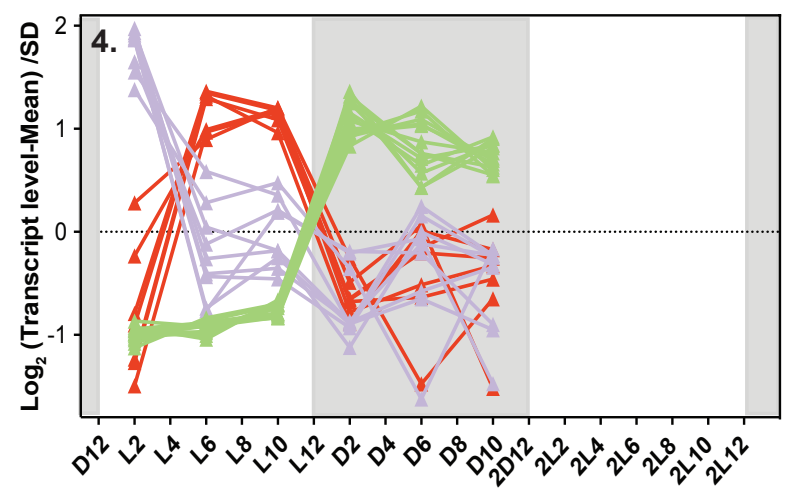




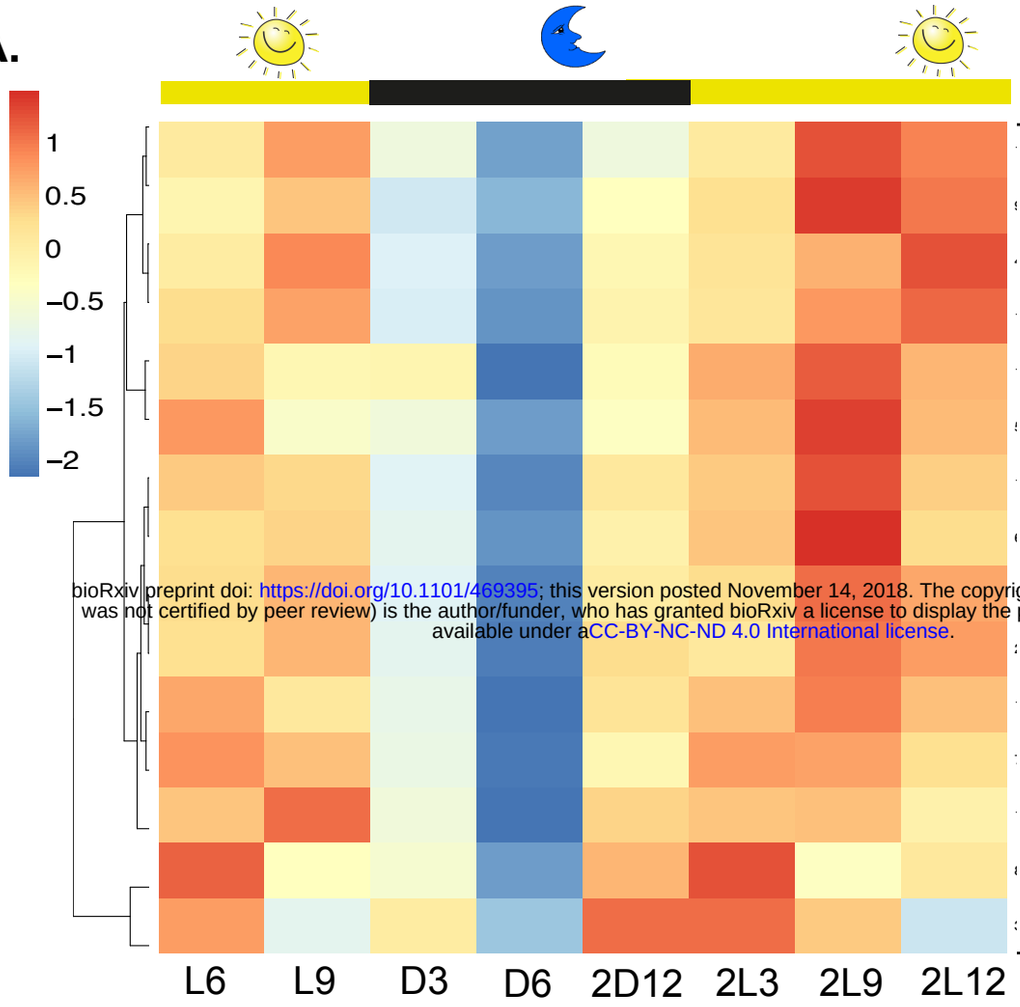
Crocosphaera watsonii WH 8501



Cyanothece ATCC 51142

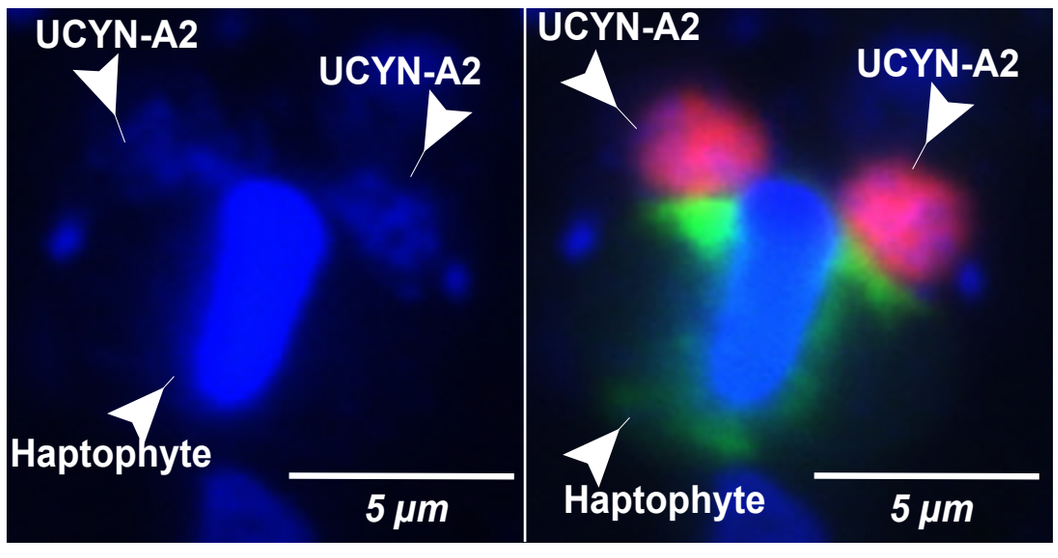


A.

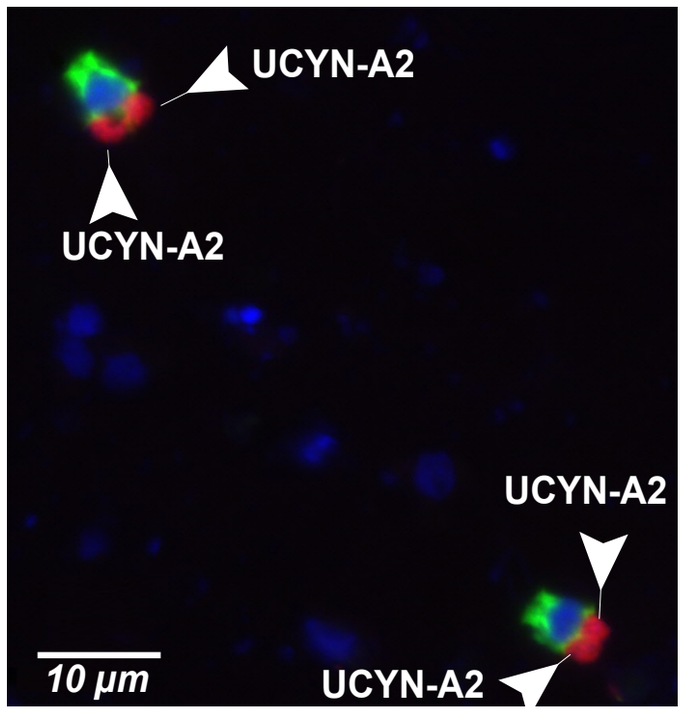


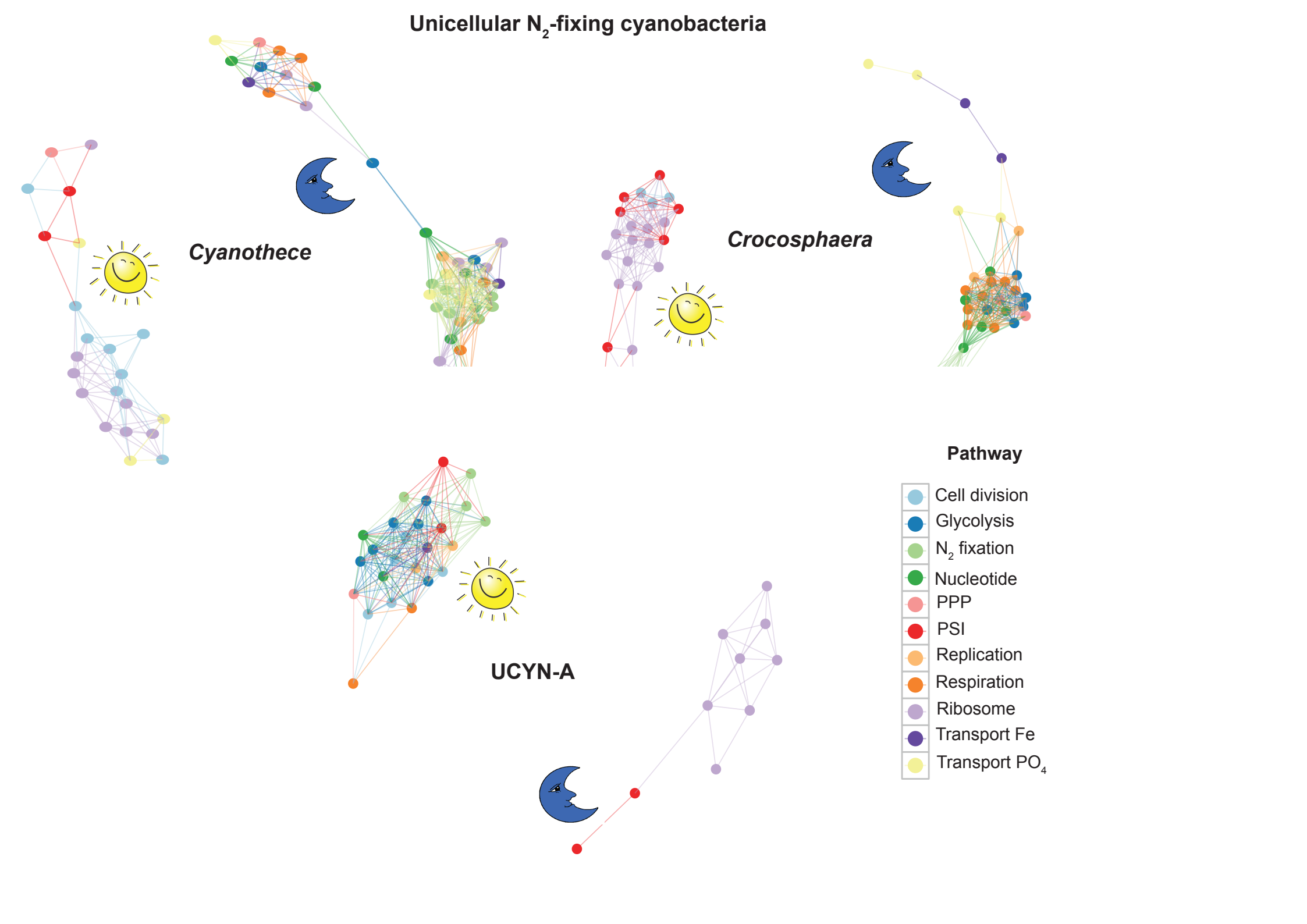
Gene_ID	Strain	Description
646530139	UCYN-A1	septum site-determining protein *minD*
646530123	UCYN-A1	DNA gyrase subunit A
2528848158	UCYN-A2	rod shape-determining protein *mreC*
2528847831	UCYN-A2	DNA polymerase III- delta prime subunit
2528847919	UCYN-A2	pyridoxal phosphate-dependent enzyme
2528847263	UCYN-A2	DNA gyrase subunit A
2528847442	UCYN-A2	cell division protein *ftsZ*
2528848139	UCYN-A2	Superfamily II DNA and RNA helicases
646530138	UCYN-A1	cell division topological specificity factor *minE*
2528847429	UCYN-A2	DNA topoisomerase I
646530091	UCYN-A1	DNA topoisomerase IV subunit A
646530302	UCYN-A1	cell division protein *ftsZ*
2528847278	UCYN-A2	cell division topological specificity factor *minE*
2528848219	UCYN-A2	Glycosyltransferases involved in cell wall biogenesis
646530433	UCYN-A1	DNA/RNA helicase, superfamily II

B.

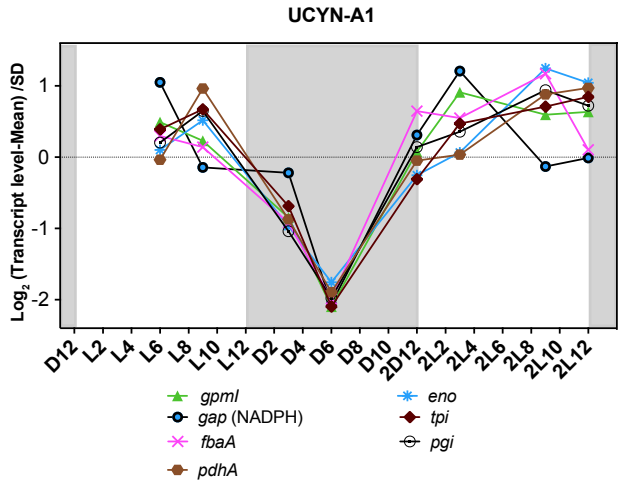


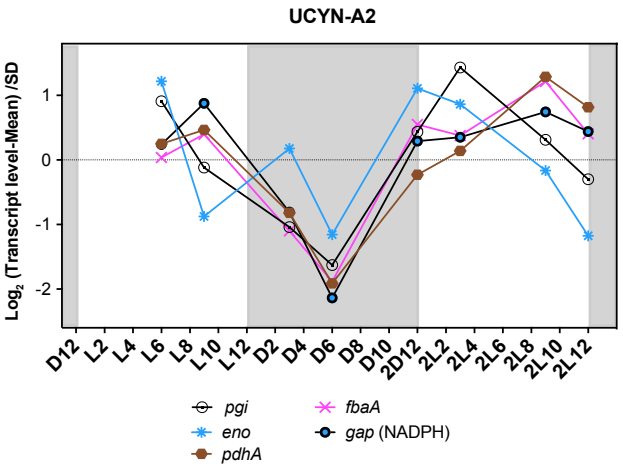
C.

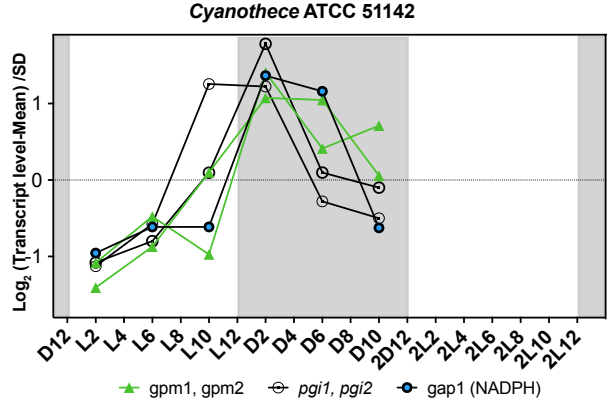


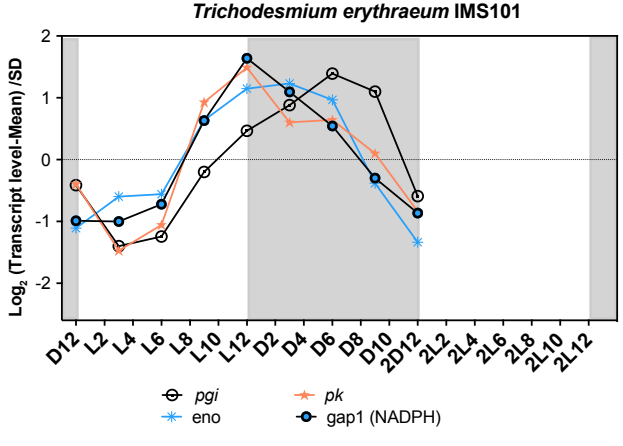

Unicellular N_2-fixing cyanobacteria
Cyanothece
Crocosphaera
UCYN-A
Pathway
Cell division
Glycolysis
N_2 fixation
Nucleotide
PPP
PSI
Replication
Respiration
Ribosome
Transport Fe
Transport PO_4

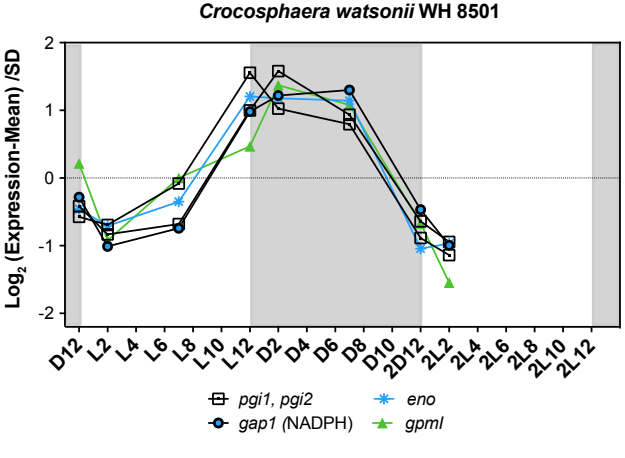
Glycolysis

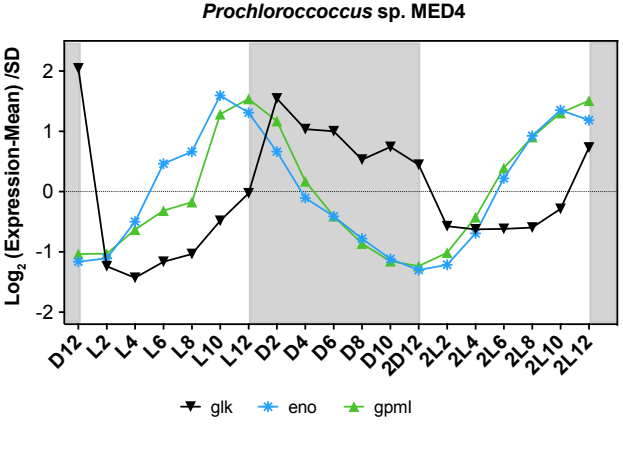

UCYN-A1
Log2 (Transcript level-Mean) /SD
gpmI
gap (NADPH)
fbaA
pdhA
eno
tpi
pgi


UCYN-A2
Log2 (Transcript level-Mean) /SD
pgi
eno
pdhA
fbaA
gap (NADPH)

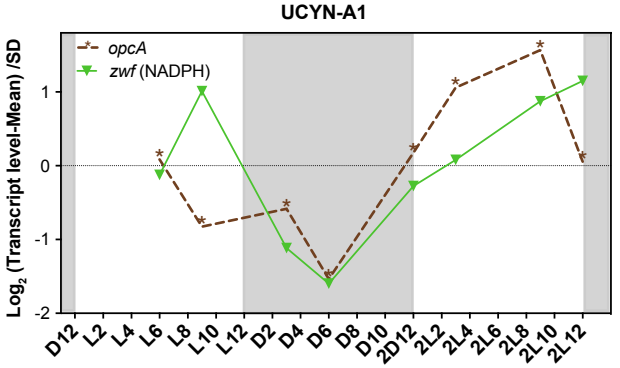

Cyanothece ATCC 51142
Log2 (Transcript level-Mean) /SD
gpm1, gpm2
pgi1, pgi2
gap1 (NADPH)

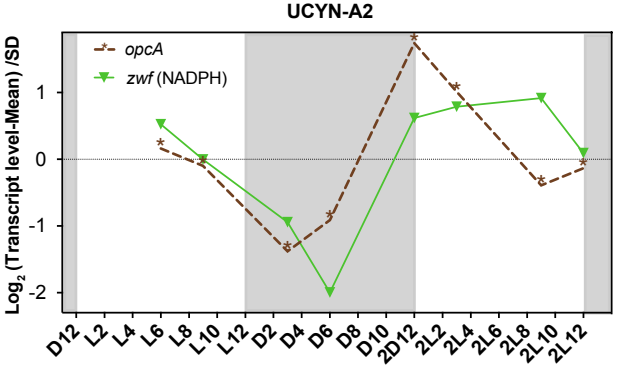

Trichodesmium erythraeum IMS101
Log2 (Transcript level-Mean) /SD
pgi
eno
pk
gap1 (NADPH)

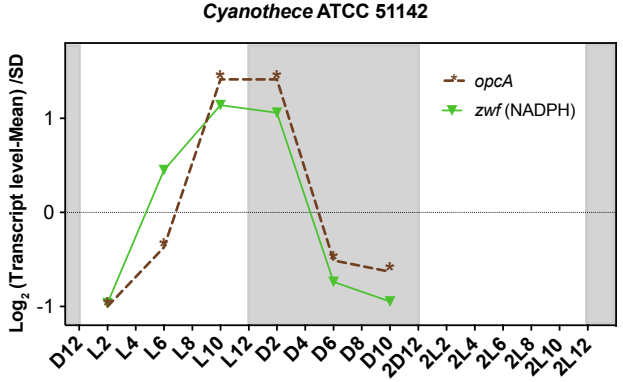

Crocosphaera watsonii WH 8501
Log2 (Expression-Mean) /SD
pgi1, pgi2
gap1 (NADPH)
eno
gpmI

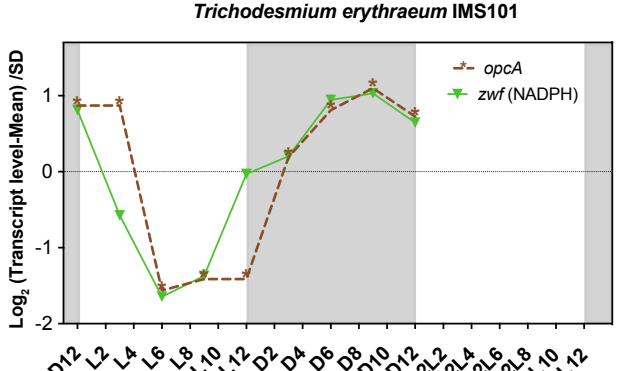

Prochloroccoccus sp. MED4
Log2 (Expression-Mean) /SD
glk
eno
gpmI

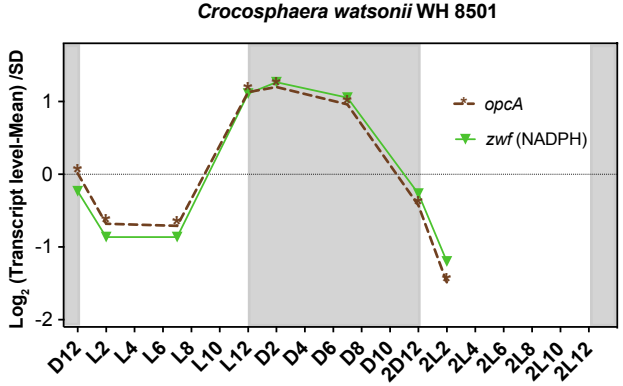
Pentose phosphate

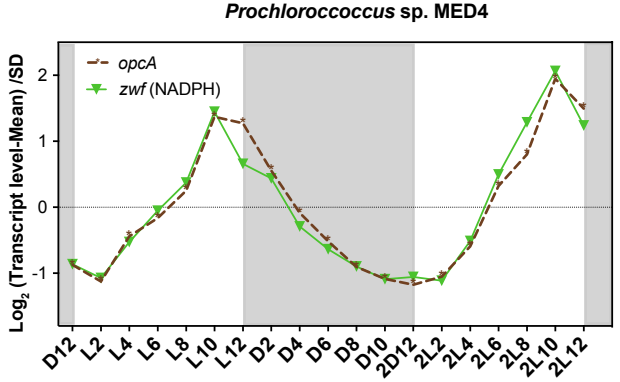


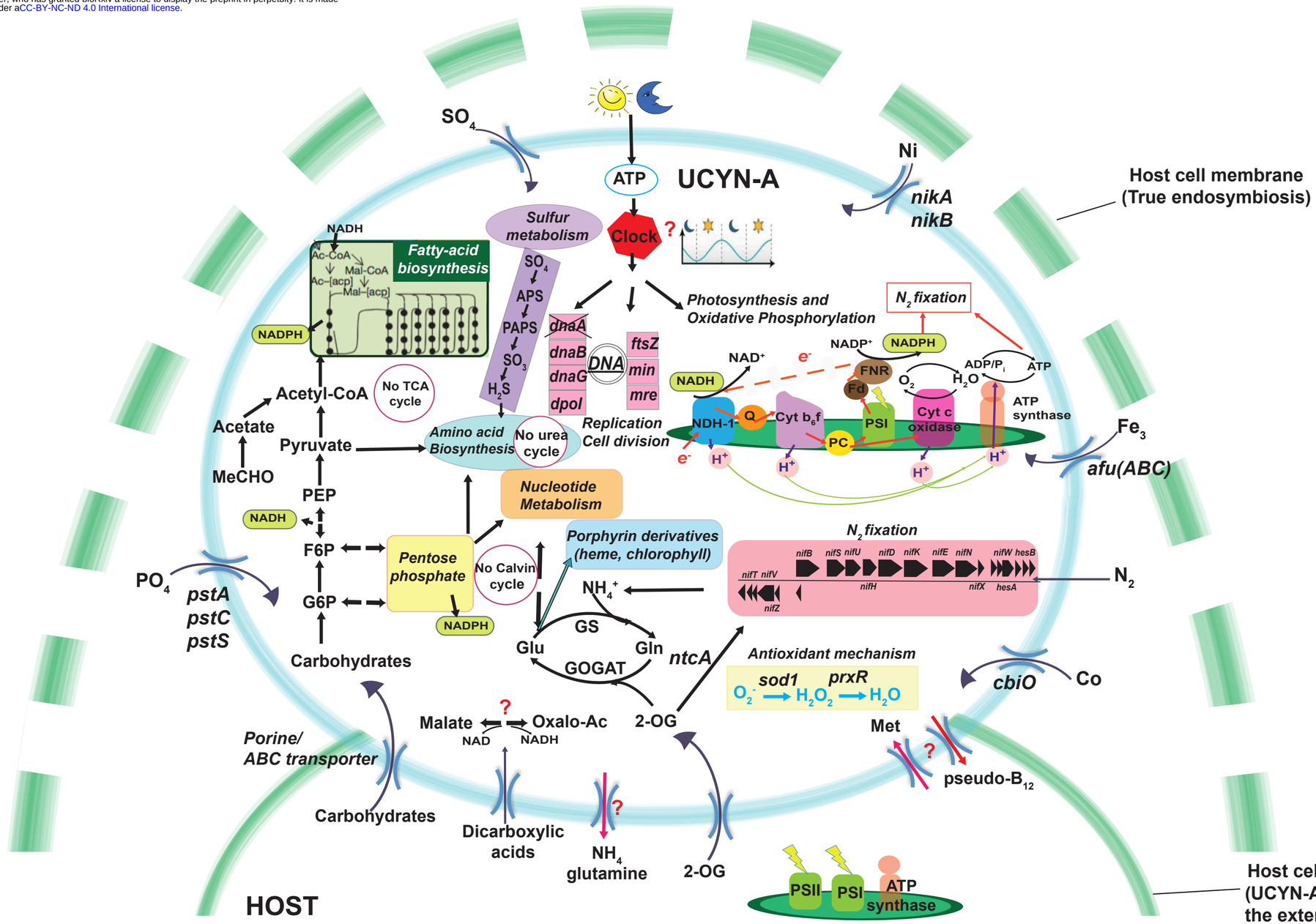


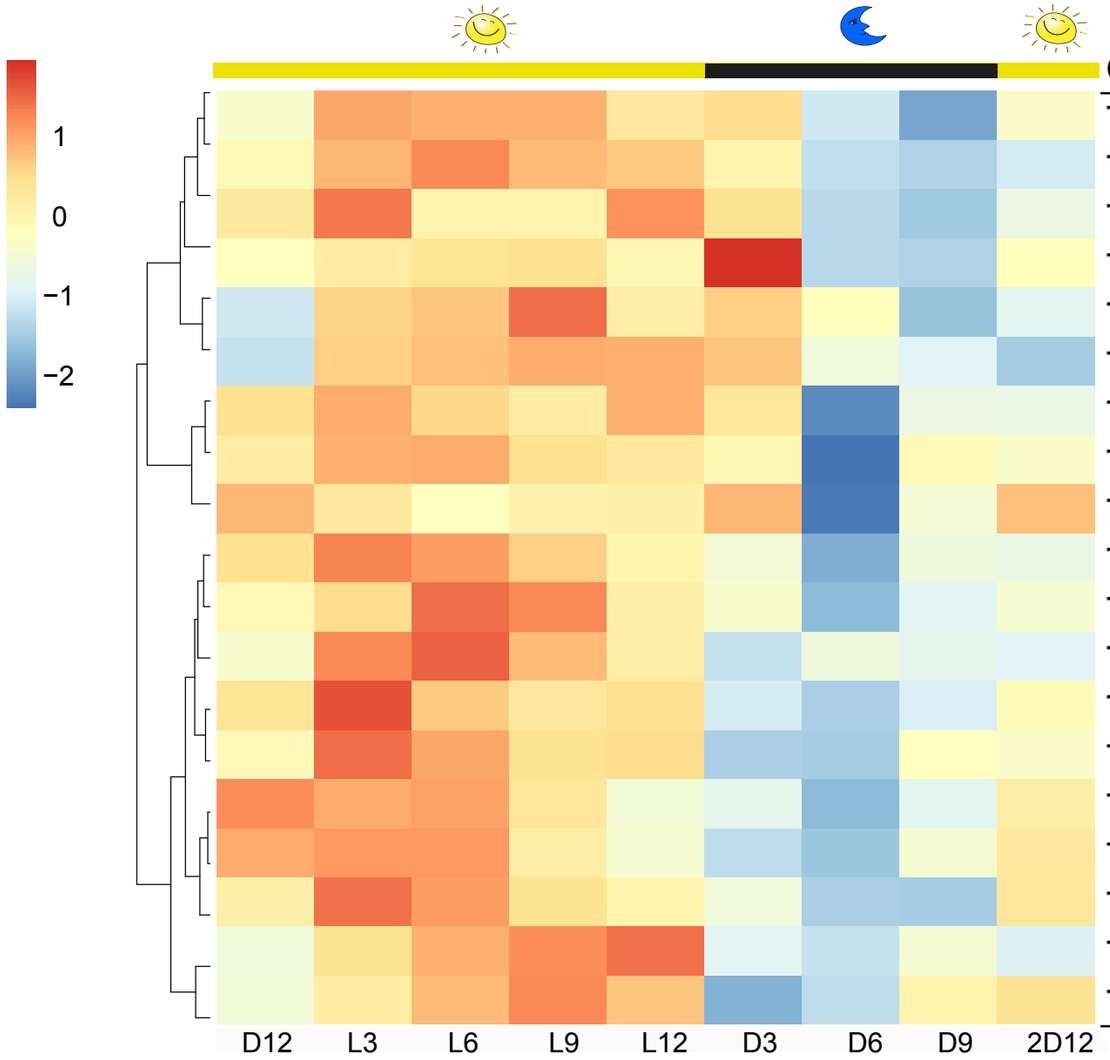






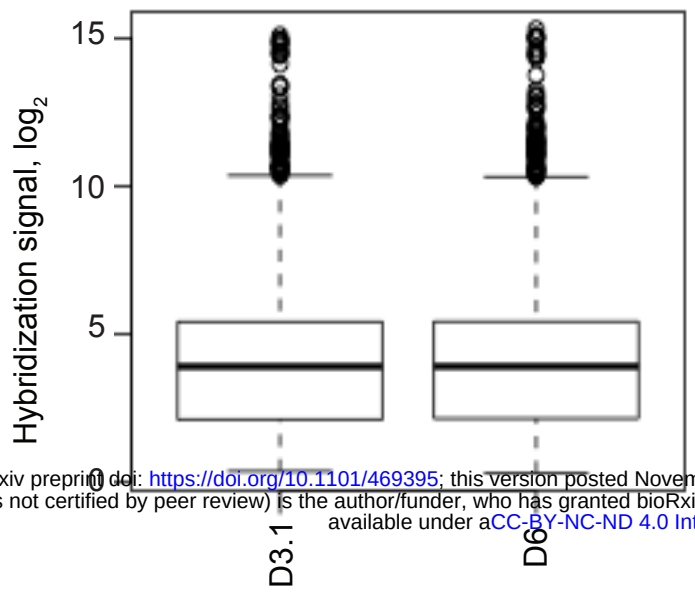




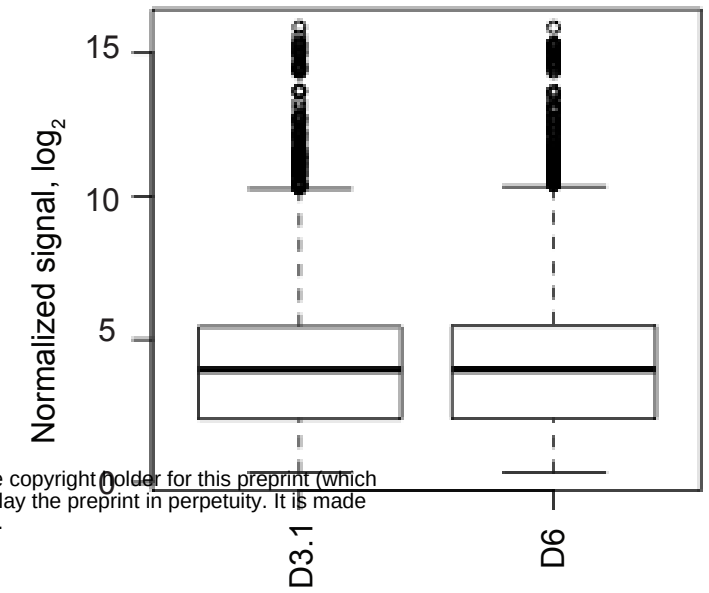


Gene_ID	Description	Pathway
Tery_3301	MutS2 family protein	Replication and Repair
Tery_2841	glucose inhibited division protein A	Cell Growth and Death
Tery_4190	single-stranded-DNA-specific exonuclease *RecJ*	Replication and Repair
Tery_0168	Holliday junction DNA helicase *RuvA*	Replication and Repair
Tery_3645	DNA mismatch repair protein *MutS*	Replication and Repair
Tery_0165	DNA recombination protein, *RuvA*	Replication and Repair
Tery_4942	cell cycle protein	Cell Growth and Death
Tery_0212	DNA polymerase III, delta subunit	Replication and Repair
Tery_2316	DNA ligase, NAD-dependent	Replication and Repair
Tery_1165	protein of unknown function DUF59	Cell Growth and Death
Tery_0163	maf protein	Cell Growth and Death
Tery_0413	phage SPO1 DNA polymerase-related protein	Replication and Repair
Tery_0001	chromosomal replication initiator protein *DnaA*	Replication and Repair
Tery_4604	DNA repair protein *RecO*	Replication and Repair
Tery_4232	ATP-dependent DNA helicase *PcrA*	Replication and Repair
Tery_0348	transposase, IS605 *OrfB* family	Replication and Repair
Tery_3302	cell cycle protein	Cell Growth and Death
Tery_3384	cell division topological specificity factor *MinE*	Cell Growth and Death
Tery_3269	endonuclease III	Replication and Repair

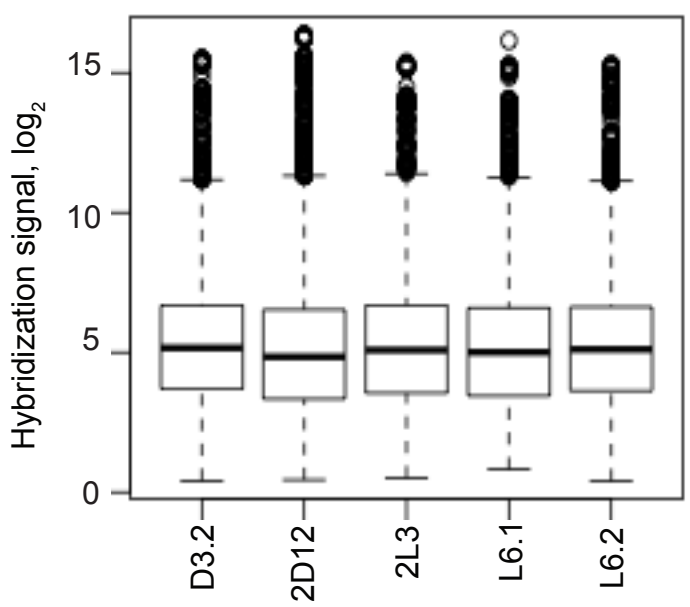
A.



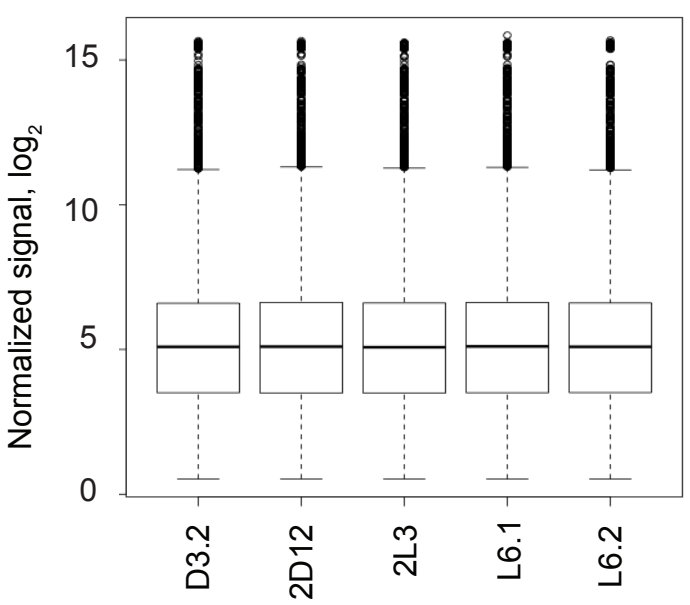
D.



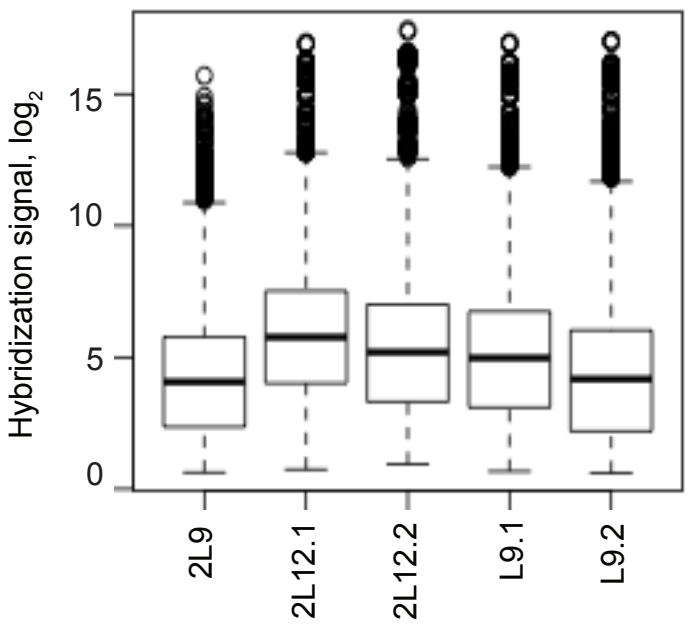
B.



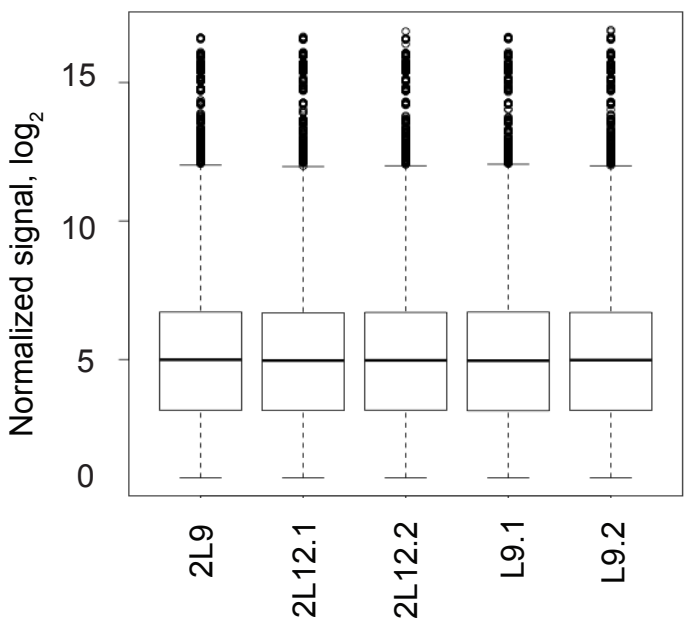
E.

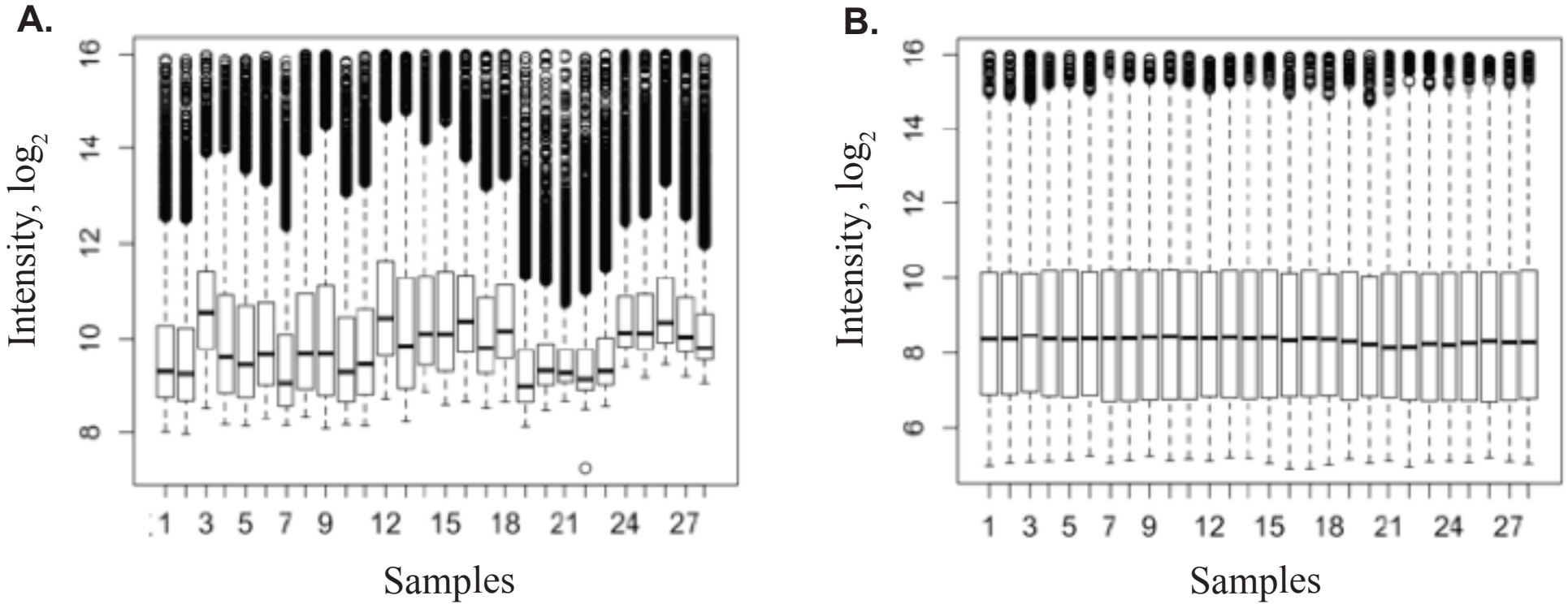


C.



F.




A.
Intensity, log2
16
14
12
10
8
1 3 5 7 9 12 15 18 21 24 27
Samples
B.
Intensity, log2
16
14
12
10
8
6
1 3 5 7 9 12 15 18 21 24 27
Samples

Number of genes	***T. erythraeum* IMS101**	***C. watsonii* WH8501**	***Cyanothece* ATCC 51142**	**UCYN-A**	***Prochlorococcus* MED4**
Targeted on the microarray	4,788	4,407	5,040	1,194	1,698
With detected transcription	4,305	4,140	4,312	365	1.511
Have diel pattern of transcription	1,620	1,708	998	316	1,370
Percent					
With detected transcription	89.9	93.9	85.6	30.6	89.0
Have diel pattern of transcription	33.8	38.8	19.8	26.5	80.7

	Sampling Time	Host associated with 1 UCYN-A2 cluster (cells ml^{-1})	Host associated with 2 UCYN-A2 clusters (cells ml^{-1})	Host associated with 2 UCYN-A2 clusters (%)
Diel cycle 1	12 pm	36.5	0.0	0.0
	3 pm	38.0	0.0	0.0
	6 pm	9.1	0.0	0.0
	9 pm	12.2	15.2	55.6
	12 am	6.1	9.1	60.0
	3 am	41.0	10.6	20.6
	6 am	7.6	0.0	0.0
	9 am	38.0	0.0	0.0
Diel cycle 2	12 pm	44.1	0.0	0.0
	3 pm	53.2	0.0	0.0
	6 pm	36.5	0.0	0.0
	9 pm	16.7	28.9	63.3
	12 am	34.9	51.7	59.6
	3 am	51.7	4.6	8.1
	6 am	66.8	0.0	0.0
	9 am	27.3	0.0	0.0

Probe Name	Target	Sequence (5' to 3')	Reference
UPRYM69	Host-A1	CACAT**A**GGAACATCCTCC	Cornejo-Castillo et al.(6)
UPRYM69 competitor	Host-A2 used as Host-A1 competitor	CACAT**T**GGAACATCCTCC	Cornejo-Castillo et al.(6)
UBRADO69	Host-A2	CACAT**T**GGAACATCCTCC	Cornejo-Castillo et al.(6)
UBRADO69 competitor	Host-A1 used as Host-A2 competitor	CACAT**A**GGAACATCCTCC	Cornejo-Castillo et al.(6)
Helper A-PRYM	*Haptophyta*	GAAAGGTGCTGAAGGAGT	Cornejo-Castillo et al.(6)
Helper B-PRYM	*Haptophyta*	AATCCCTAGTCGGCATGG	Cornejo-Castillo et al.(6)
UCYN-A1 732	UCYN-A1	GTT**A**CGGTCCAGTAGCAC	Krupke et al.(5)
UCYN-A1 competitor	UCYN-A2 used as UCYN-A1 competitor	GTT**G**CGGTCCAGTAGCAC	Cornejo-Castillo et al.(6)
UCYN-A2 732	UCYN-A2	GTT**G**CGGTCCAGTAGCAC	Cornejo-Castillo et al.(6)
UCYN-A2 competitor	UCYN-A1 used as UCYN-A2 competitor	GTT**A**CGGTCCAGTAGCAC	Krupke et al.(5)
Helper A-732	UCYN-A	GCCTTCGCCACCGATGTTCTT	Krupke et al.(5)
Helper B-732	UCYN-A	AGCTTTCGTCCCTGAGTGTCA	Krupke et al.(5)