Supplemental Results

Cyclic genes are more conserved than acyclic genes in Cyanothece 51142

An interesting observation which is not central to our study is that those genes found to be cyclic (either diurnal or circadian) using Haystack (6) are more likely to be conserved than non-cyclic genes. The threshold chosen for network generation significantly affects its topology, and may impact the observed relationship between centrality and conservation. We therefore varied the correlation threshold used for network generation and found that enrichment of bottlenecks in homologs for the various organisms was significant and observed over a range of correlation values between 0.7-0.95.

Bottlenecks in wreath networks are more conserved than hubs

Previous studies have found that the importance of bottlenecks was most affected by their connectedness, i.e. that bottlenecks that are also hubs are the most important (7). We therefore analyzed the enrichment of bottlenecks, hubs and combinations of the two in the wreath network at the indicated threshold. Contrary to previous findings, the bottlenecks and bottleneck-nonhub genes were the most enriched in conserved genes, and they were the only groups that where the enrichment was significant. This indicates that bottlenecks and hubs are playing very different roles in these kinds of networks.

Clustering and cross-validation in predictive model construction

Previously, biclustering methods and additional data were used to define co-regulated clusters (8). For our application hierarchical clustering should be appropriate since there are a much smaller set of different conditions and these conditions are nearly all related to cyclic processes. As described in Materials and Methods and Figure 2 the approach was validated by defining four sets of non-similar time series from the complete data set, then constructing models based on leaving each set out of the training data, and finally predicting the behavior of all target clusters for the data that was left out. This process is meant to ensure that the model is not being over trained and that it will generalize to other data sets that are not part of the training set. The performance of the model was evaluated by predicting the behavior of each target then comparing the predicted and observed expression levels using either Pearson correlation or a normalized root mean

square deviation (nRMSD) measure. Overall model performance was calculated as the mean of the performance measures for each target weighted by the number of genes represented in each target, a gene-normalized performance measure.

Inference of regulatory networks in Cyanothece 51142

We used two related but distinct methods for inference of regulatory networks from transcriptional data. The first is using the machine learning approach we have outlined in the main text (Figure 2). In this approach ordinary differential equations (ODEs) are used to relate the expression levels of a parsimonious set of regulators to the expression level of the target, either a coexpressed cluster or another regulator. In this process regulators with expression profiles that are highly correlated ($R > 0.9$) with the target are excluded from consideration as potential influences. This filter ensures that the inferred regulatory influences are predictive of behavior and not merely highly correlated. This approach also provides directionality of influence (from regulators to target) and a weight of individual influences (see the β_j terms in equations 1 and 2 below). The second approach, the context likelihood of relatedness (CLR) method, relies on mutual information to determine the shared information (represented by similarity or difference) between the expression profiles of two regulators. A filter is applied to each regulator to select those relationships, which have the highest mutual information relative the mutual information scores between that regulator and all other regulators. This filter provides a method to discriminate direct from indirect regulatory relationships but does not determine directionality of relationships (the mutual information of A to B will be identical to that of B to A) nor a weight that describes the contribution of one expression level to the other. The CLR score is an indication of shared information, not causal influence. We have previously used CLR to infer regulatory relationships from high-throughput data and validated key regulatory relationships identified (9).

To predict a regulatory network involved in cyclic processes in *Cyanothece* 51142 we first applied the least angle regression method using the set of transcriptional regulators as both the pool of potential regulators and targets for inference using a cross-validation approach identical to the approach shown in Figure 2 except that the targets are also

regulators. We filtered the resulting regulatory network to include only relationships with larger weights (> 0.1) and combined it with the network described in the main text that relates regulators with target coexpressed clusters. The resulting network is shown in Figure S2.

We similarly applied the CLR method to the expression profiles of all transcription factors across the three cyclic experiments used and show the relationships with CLR Z scores of 4 or higher. The entire CLR-inferred regulatory network is shown in Figure S3A. In Fig. S3B we show a subnetwork of the CLR network focused on the *rpaA*, *patB*, and *ntcA* regulators and show the corresponding subnetwork of the least angle regression network in Figure S3C. Interestingly, the two methods of network inference show very similar network structures, especially for this important subnetwork.

References

1. Liang J, Scappino L, & Haselkorn R (1993) The patB gene product, required for growth of the cyanobacterium Anabaena sp. strain PCC 7120 under nitrogen-limiting conditions, contains ferredoxin and helix-turn-helix domains. *Journal of bacteriology* 175(6):1697-1704.
2. Summerfield TC & Sherman LA (2007) Role of sigma factors in controlling global gene expression in light/dark transitions in the cyanobacterium Synechocystis sp. strain PCC 6803. *Journal of bacteriology* 189(21):7829-7840.
3. Takai N, *et al.* (2006) 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.
4. Summers ML, Wallis JG, Campbell EL, & Meeks JC (1995) Genetic evidence of a major role for glucose-6-phosphate dehydrogenase in nitrogen fixation and dark growth of the cyanobacterium Nostoc sp. strain ATCC 29133. *Journal of bacteriology* 177(21):6184-6194.
5. Singh AK, Li H, Bono L, & Sherman LA (2005) Novel adaptive responses revealed by transcription profiling of a Synechocystis sp. PCC 6803 delta-isiA mutant in the presence and absence of hydrogen peroxide. *Photosynthesis research* 84(1-3):65-70.
6. Michael TP, *et al.* (2008) Network Discovery Pipeline Elucidates Conserved Time-of-Day-Specific cis-Regulatory Modules. *PLoS genetics* 4(2):e14.

7. Yu H, Kim PM, Sprecher E, Trifonov V, & Gerstein M (2007) The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics. *PLoS Comput Biol* 3(4):e59.
8. Bonneau R*, et al.* (2006) The Inferelator: an algorithm for learning parsimonious regulatory networks from systems-biology data sets de novo. *Genome biology* 7(5):R36.
9. Yoon H, McDermott JE, Porwollik S, McClelland M, & Heffron F (2009) Coordinated regulation of virulence during systemic infection of Salmonella enterica serovar Typhimurium. *PLoS Pathog* 5(2):e1000306.
10. Colon-Lopez MS, Sherman DM, & Sherman LA (1997) Transcriptional and translational regulation of nitrogenase in light-dark- and continuous-light-grown cultures of the unicellular cyanobacterium Cyanothece sp. strain ATCC 51142. *Journal of bacteriology* 179(13):4319-4327.
11. Li H, Sherman DM, Bao S, & Sherman LA (2001) Pattern of cyanophycin accumulation in nitrogen-fixing and non-nitrogen-fixing cyanobacteria. *Arch Microbiol* 176(1-2):9-18.
12. Schneegurt MA, Sherman DM, & Sherman LA (1997) Composition of the carbohydrate granules of the cyanobacterium, Cyanothece sp. strain ATCC 51142. *Arch Microbiol* 167(2/3):89-98.
13. Kanehisa M, Goto S, Kawashima S, & Nakaya A (2002) The KEGG databases at GenomeNet. *Nucleic Acids Res* 30(1):42-46.
14. Apweiler R*, et al.* (2000) InterPro--an integrated documentation resource for protein families, domains and functional sites. *Bioinformatics* 16(12):1145-1150.
15. McDermott J, Guerquin M, Frazier Z, Chang AN, & Samudrala R (2005) BIOVERSE: enhancements to the framework for structural, functional and contextual modeling of proteins and proteomes. *Nucleic Acids Res* 33(Web Server issue):W324-325.
16. Efron B, Johnstone I, Hastie T, & Tibshirani R (2003) Least angle regression. *Annals of Statistics* 32:407-499.
17. Welsh EA*, et al.* (2008) The genome of Cyanothece 51142, a unicellular diazotrophic cyanobacterium important in the marine nitrogen cycle. *Proceedings of the National Academy of Sciences of the United States of America* 105(39):15094-15099.
18. Faith JJ*, et al.* (2007) Large-scale mapping and validation of Escherichia coli transcriptional regulation from a compendium of expression profiles. *PLoS biology* 5(1):e8.

Figure S1. Topological enrichment of co-expression networks. Topological bottleneck genes were identified from the co-expression wreath network of *Cyanothece* 51142 under 12 h LD cycles by calculating topological betweenness. The fold enrichment of homologous genes in *Escherichia coli*, *Synechocystis* sp. PCC 6803, *Arabidopsis thaliana*, and those that are homologs in all members of a set of organisms that are photosynthesizers (*Synechocystis*, *A. thaliana*, and *Anabaena* sp. PCC 7120) was calculated as the ratio between the percentage of bottleneck genes that have a homolog(s) in that species and the percentage of non-bottlenecks that have a homolog(s). All results shown are statistically significant using a Chi-square test and the bottleneck enrichment observed is statistically significant relative when compared to general cyclic genes. These results show that cyclic genes and topological bottlenecks are significantly more likely to be present in several organisms, including *A. thaliana*, than are non-cyclic or non-bottleneck genes.

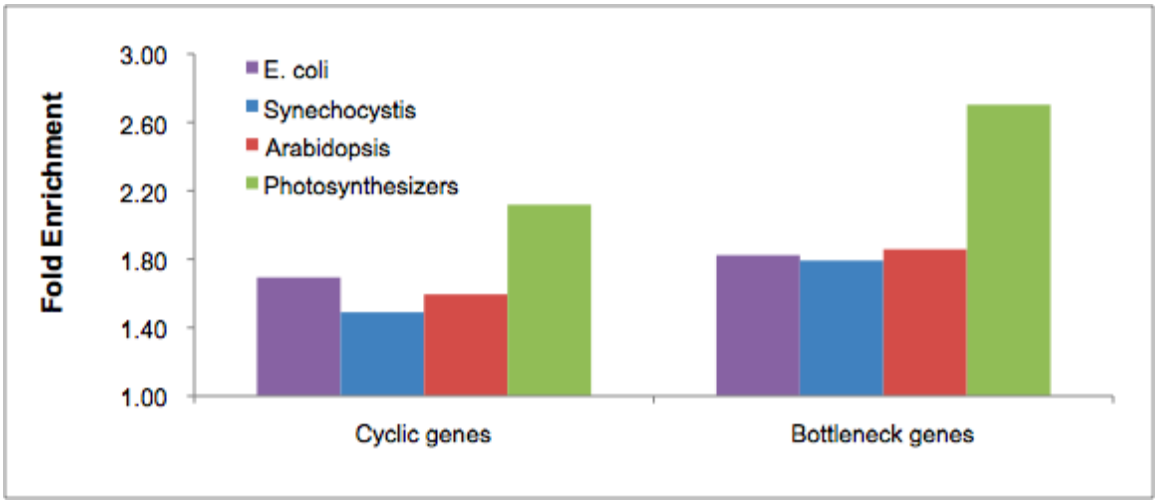
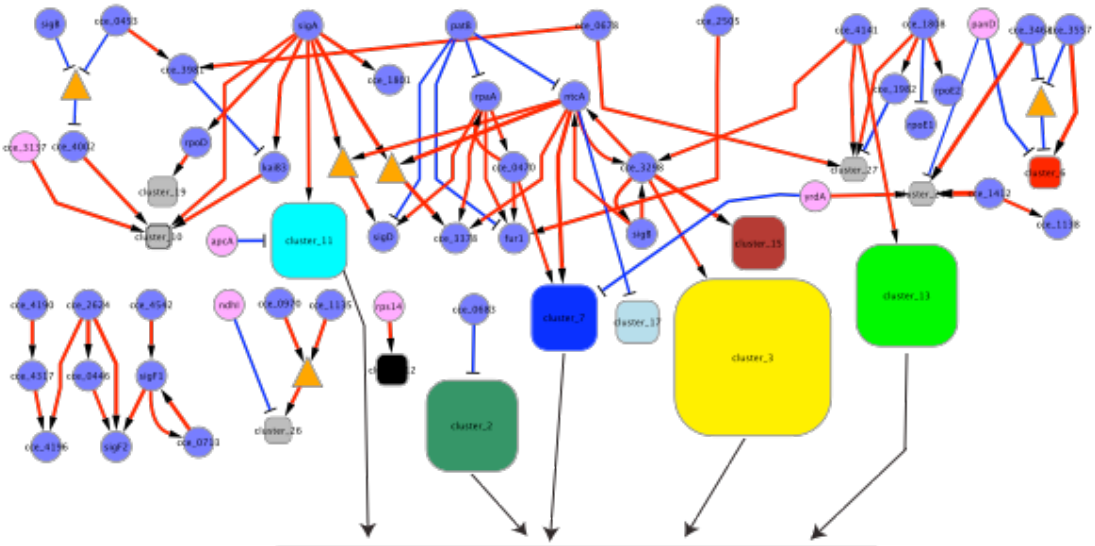


Figure S2. Transcriptional regulatory model of cyclic behavior in *Cyanothece* 51142. We constructed a model of predicted causal regulatory influences based on regulators (transcription/translation factors [blue circles] and bottleneck genes [pink circles]) and co-expressed clusters (squares). The network shows all regulatory influences on targets (regulators or clusters) that could be predicted with reasonable accuracy (R >= 0.5). Red arrows represent positive regulation and blue lines represent negative regulation. Orange triangles represent logical combinations of regulatory influences predicted by the model.



Label	Description	Pattern
cluster_6 (15):	psbA4	D5-D9
cluster_13 (103):	apc, cpc, rbc, ccm, atp, glgA1, glgB1, glgB2	L1
cluster_24 (4):	hupS	D1-D9
cluster_27 (4):	petF3, petF4, psb27	L1
cluster_3 (173):	PSII, ntcA, sigB, ftsH1, ftsH2, ndhEHI	L5
cluster_15 (42):	psbC, hypE, rpoE1, ftsW	L1-L5
cluster_17 (28):	feoA2, feoB2, hupL, patB, aphA	D1
cluster_7 (57):	PSI, rpaA, cmp, sigD	L5-L9
cluster_2 (90):	kaiB4, mazE, sigF1, sigF2	L1
cluster_12 (14):	ribosome, groES, groEL, dnaK, ho2, hemN	D5
cluster_26 (7):	unknown	D5-D9
cluster_11 (70):	kaiB3, rpoD, lrtA, glnB, coxABC, hypA	L9-D1
cluster_10 (10):	glgP1	L9-D5
cluster_19 (11):	pnt, gnd	L9-D1

Figure S3. Regulatory network inferred by the context likelihood of relatedness method. A. A network was inferred between regulators and several important functional complexes used in the study using the Context Likelihood of Relatedness (CLR) method. Red edges indicate positive regulation and blue edges indicate negative regulation. The width of the edge indicates relative strength of the inferred association. Directionality is not predicted using CLR. **B.** A subnetwork from the CLR network (panel A) focusing on several important regulators. **C.** The same subnetwork from the Inferelator network (Figure 5) that shows many of the same relationships. The Inferelator predicts directionality to associations (influences). These results indicate that the two different network inference methods reveal a similar regulatory structure.

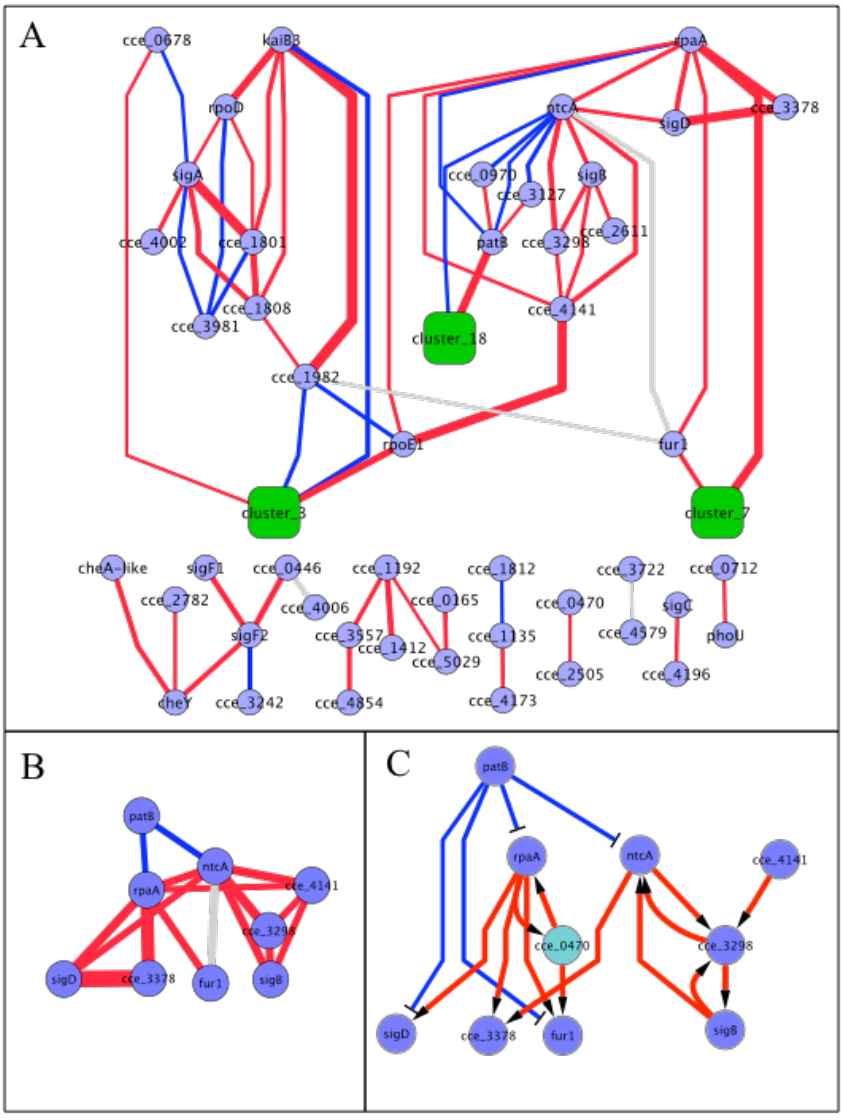


Figure S4. CO_2 uptake under 12 h LD conditions. CO_2 uptake was assayed every 4 hours over 48 hours in a 12 hour light/dark cycle. Error bars indicate the standard deviation from two biological replicates.

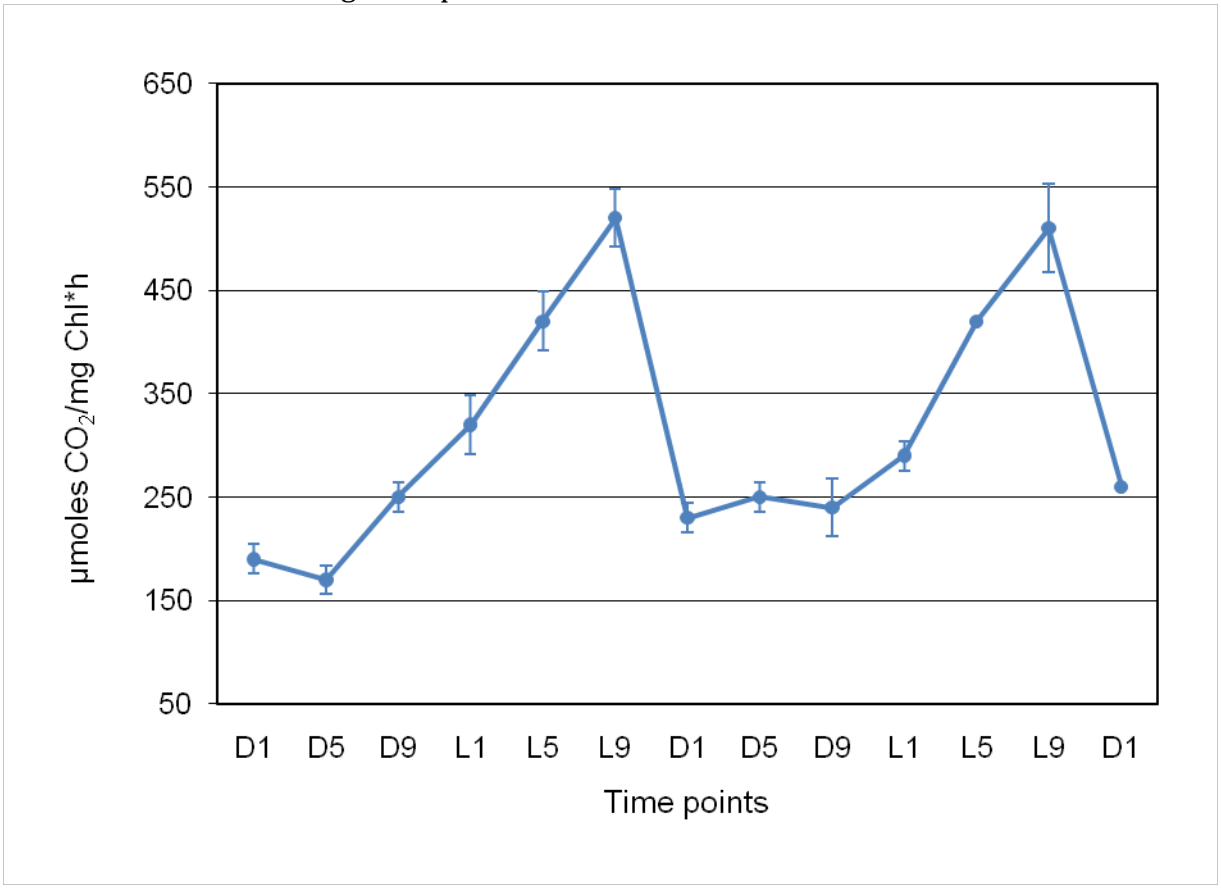


Table S1. Functions enriched in topological bottlenecks. PCC network: co-expression network determined using Pearson correlation and a threshold of 0.9125. CLR network: network inferred using the Context Likelihood of Relatedness method. N: number of bottleneck genes with the indicated annotation. P-value: calculated using Fisher's exact test against the remainder of the non-bottleneck genes in the network. Grey background indicates that the relationship is not significant.

	PCC Network			CLR Network		
Function	**Fold Enrichment**	**P value**	**N**	**Fold Enrichment**	**P value**	**N**
KEGG: Photosynthesis	2.17	1.8E-02	13	2.06	2.7E-02	13
Nitrogen fixation, nitrogenase	*0.12*	*1.1E-02*	*1*	*0.00*	2.8E-03	*0*
KEGG: Oxidative phosphorylation	2.02	1.2E-01	7	2.55	2.2E-02	9
KEGG: Carbon fixation	2.88	5.6E-02	5	3.39	1.9E-02	6
KEGG: Pentose phosphate pathway	2.24	1.3E-01	5	2.97	3.3E-02	6
Circadian regulated genes	0.88	2.8E-01	49	0.75	2.6E-02	48
GO: Proton-transporting two-sector ATPase complex	6.00	1.4E-03	6	2.64	1.2E-01	4
KEGG: ATP synthesis coupled proton transport	5.00	7.2E-03	5	1.98	3.2E-01	3
GO: Iron ion binding	1.61	3.1E-01	6	2.64	1.2E-02	10
GO: Two-component response regulator activity	3.43	1.7E-02	6	1.52	4.2E-01	5
GO: DNA binding	3.08	4.1E-03	10	1.32	4.6E-01	9
GO: Carbohydrate metabolism	4.67	2.1E-03	7	1.32	6.2E-01	4
KEGG: Starch and sucrose metabolism	5.33	1.4E-02	4	0.49	5.0E-01	1
GO: Catalytic activity	3.14	2.2E-03	11	0.88	7.8E-01	6
GO: Electron transport	0.51	1.0E-01	6	1.81	1.8E-02	21
GO: Cytochrome-c oxidase activity	0.00	3.9E-01	0	15.84	8.2E-04	4
GO: Mannitol transport	0.00	3.2E-01	0	-	5.6E-04	3
GO: Response to stress	0.00	3.9E-01	0	11.88	6.1E-03	3

Table S2. Regulators identified in *Cyanothece* 51142

ORF	Name	Description	Cyclicity
cce_0012			
cce_0115	*cce_0115*	response regulator	acyclic
cce_0145	*kaiB4*	putative circadian clock protein; KaiB4	acyclic
cce_0165	*cce_0165*	two-component response regulator	acyclic
cce_0198	*ntcB*	LysR family transcriptional regulatory protein, nitrogen assimilation transcriptional activator; NtcB	circadian
cce_0289	*cce_0289*	two-component response regulator	diurnal
cce_0298	*rpaA*	two-component response regulator; RpaA	diurnal
cce_0310	*aslB*	putative arylsulfatase regulatory protein; AslB	diurnal
cce_0350	*cce_0350*	hypothetical protein	acyclic
cce_0391	*cce_0391*	protein containing a Ribonucleotide reductase regulator NrdR-like domain	diurnal
cce_0422	*kaiC1*	circadian clock protein; KaiC1	insignificant
cce_0423	*kaiB1*	circadian clock protein; KaiB1	diurnal
cce_0424	*kaiA*	circadian clock protein; KaiA	diurnal
cce_0435	*kaiB3*	circadian clock protein; KaiB3	circadian
cce_0446	*cce_0446*	two-component response regulator	insignificant
cce_0448	*cce_0448*	response regulator	diurnal
cce_0452	*cce_0452*	protein containing a transcriptional regulator, Rrf2 domain	diurnal
cce_0453	*abrB*	putative transcriptional regulator AbrB	circadian
cce_0461	*ntcA*	nitrogen-responsive regulatory protein; NtcA	diurnal
cce_0470	*cce_0470*	protein containing an Anti-sigma factor antagonist domain	insignificant
cce_0595	*glk*	glucokinase; Glk	insignificant
cce_0601	*rpoD*	RNA polymerase sigma factor; RpoD	diurnal
cce_0644	*sigB*	RNA polymerase sigma factor; SigB	diurnal
cce_0656	*cce_0656*	two-component hybrid sensor and regulator	circadian
cce_0657	*cce_0657*	two-component response regulator	diurnal
cce_0678	*cce_0678*	two-component response regulator	diurnal
cce_0683	*cce_0683*	putative transcriptional regulator	insignificant
cce_0712	*cce_0712*	two-component response regulator	diurnal
cce_0713	*cce_0713*	two-component response regulator	acyclic
cce_0716	*cce_0716*	two-component hybrid sensor and regulator	circadian
cce_0754			
cce_0817	*cce_0817*	two component transcriptional regulator	diurnal
cce_0875	*sigA*	RNA polymerase sigma factor; SigA	diurnal
cce_0970	*cce_0970*	two-component transcription regulator	insignificant
cce_0974	*cce_0974*	transcriptional Regulator, TetR family	insignificant
cce_1135	*cce_1135*	two-component response regulator	circadian
cce_1138	*cce_1138*	two-component hybrid sensor and regulator	insignificant
cce_1171	*cce_1171*	protein containing a TPR-like helical domain	acyclic
cce_1185	*cce_1185*	two-component hybrid sensor and regulator	diurnal
cce_1186	*cce_1186*	two-component hybrid sensor and regulator	diurnal
cce_1192	*cce_1192*	protein containing winged helix repressor DNA-binding and a helix-turn-helix, type 11 domains	acyclic

cce_1379	*exsB*	transcription regulator; ExsB	acyclic
cce_1388	*hrcA*	heat-inducible transcriptional repressor; HrcA	circadian
cce_1412	*cce_1412*	two-component response regulator	insignificant
cce_1423	*cce_1423*	transcriptional regulator	insignificant
cce_1498	*cce_1498*	nitrogen assimilation regulatory protein, cyclic nucleotide binding, sigma-54 factor interacting region	insignificant
cce_1520	*cce_1520*	two-component response regulator	diurnal
cce_1521	*cce_1521*	protein containing PAS, GGDEF, EAL, Response regulator receiver, CheY-like, and PAS fold-3 domains	insignificant
cce_1630	*cce_1630*	two component transcriptional regulator, winged helix family	circadian
cce_1695	*cce_1695*	two-component response regulator	diurnal
cce_1699	*cce_1699*	transcriptional regulator, MerR family	acyclic
cce_1725	*cce_1725*	two-component transcriptional regulatory protein	insignificant
cce_1751	*sasA*	adaptive-response sensory histidine kinase; SasA	diurnal
cce_1768	*cce_1768*	probable transcriptional regulator	diurnal
cce_1777	*cce_1777*	probable two-component hybrid sensor and regulator	diurnal
cce_1801	*cce_1801*	protein containing an Anti-sigma factor antagonist domain	circadian
cce_1808	*cce_1808*	probable two-component system response regulator, OmpR subfamily	insignificant
cce_1812	*cce_1812*	probable transcriptional regulator	diurnal
cce_1873	*sigC*	RNA polymerase sigma factor; SigC	diurnal
cce_1877	*cce_1877*	protein containing a PAS domain, GGDEF domain, EAL domain, Response regulator receiver domain, CheY-like domain, and a PAS fold domain	insignificant
cce_1898	*patB*	probable transcriptional regulator; PatB	circadian
cce_1951	*fur2*	probable ferric uptake regulator; Fur2	circadian
cce_1952	*cce_1952*	two-component response regulator receiver protein containing a CheY-like domain	diurnal
cce_1982	*cce_1982*	two-component response regulator CheY subfamily	circadian
cce_2304	*cce_2304*	protein containing regions 3 and 4 of a tetratricopeptide-like helical and sigma factor	insignificant
cce_2330	*fur1*	ferric uptake regulation protein; Fur1	diurnal
cce_2337	*sigB*	putative RNA polymerase sigma factor; SigB	acyclic
cce_2365	*cce_2365*	two-component response regulator containing a luxR domain	diurnal
cce_2375	*cce_2375*	two-component hybrid sensor and regulator	insignificant
cce_2376	*cce_2376*	probable diguanylate cyclase (GGDEF) domain containing two-component response regulator	
cce_2424	*rpoE2*	putative RNA polymerase sigma-E factor; RpoE2	insignificant
cce_2505	*cce_2505*	two-component hybrid sensor and regulator	insignificant
cce_2507	*cce_2507*	two-component response regulator	circadian
cce_2508	*cce_2508*	two-component hybrid sensor and regulator	acyclic
cce_2509	*cce_2509*	two-component hybrid sensor and regulator	acyclic
cce_2522	*cce_2522*	ROK family protein	acyclic

cce_2566	*sigF1*	RNA polymerase sigma factor; sigF1	diurnal
cce_2611	*cce_2611*	protein containing PAS, GGDEF, Response regulator receiver, and GAF domains	diurnal
cce_2624	*cce_2624*	CheA-like two-component hybrid sensor and regulator	diurnal
cce_2782	*cce_2782*	LysR family transcriptional regulator	diurnal
cce_2836	*cce_2836*	putative transcriptional regulator	insignificant
cce_2881	*fur3*	ferric uptake regulation protein; Fur3	diurnal
cce_3127	*cce_3127*	transcription regulator, Fur family	insignificant
cce_3174	*cce_3174*	transcriptional regulator	acyclic
cce_3190	*cce_3190*	two-component response regulator	diurnal
cce_3242	*cce_3242*	protein containing a transcriptional regulator AbrB domain	acyclic
cce_3298	*cce_3298*	conserved unknown protein	circadian
cce_3320	*cce_3320*	two-component response regulator	insignificant
cce_3321	*cce_3321*	anti-sigma factor antagonist	insignificant
cce_3375	*cce_3375*	transcriptional regulator	acyclic
cce_3378	*cce_3378*	two-component response regulator	diurnal
cce_3468	*cce_3468*	transcriptional regulator, ArsR family	insignificant
cce_3519	*phoU*	phosphate uptake regulator; PhoU	diurnal
cce_3556	*cce_3556*	transcriptional regulator, ArsR family	acyclic
cce_3557	*cce_3557*	two-component response regulator	acyclic
cce_3558	*cce_3558*	two-component hybrid sensor and regulator	
cce_3559	*cce_3559*	two-component response regulator	diurnal
cce_3586	*sigF2*	RNA polymerase sigma factor; SigF2	diurnal
cce_3594	*sigD*	RNA polymerase sigma factor; SigD	circadian
cce_3714	*cce_3714*	putative two-component system response regulator	diurnal
cce_3715	*cce_3715*	protein containing PAS, GGDEF, Response regulator receiver, and GAF domains	diurnal
cce_3722	*cce_3722*	two-component hybrid sensor and regulator	diurnal
cce_3723			
cce_3731	*rbcR*	putative Rubisco transcriptional regulator, LysR family; RbcR	diurnal
cce_3753	*cce_3753*	transcriptional regulatory protein, LysR family	insignificant
cce_3796	*cce_3796*	transcriptional regulator, AraC family	diurnal
cce_3890	*cce_3890*	Probable two-component response regulator	diurnal
cce_3895	*cce_3895*	two-component response regulator	diurnal
cce_3937	*cce_3937*	two-component hybrid sensor and regulator	circadian
cce_3938	*cce_3938*	two-component response regulator	acyclic
cce_3960			
cce_3981	*cce_3981*	transcriptional regulator, ArsR family	diurnal
cce_4002	*cce_4002*	two-component response regulator	diurnal
cce_4006	*cce_4006*	two-component sensor histidine kinase	insignificant
cce_4141	*cce_4141*	conserved hypothetical protein	diurnal
cce_4142	*rpoE1*	RNA polymerase sigma-E factor; RpoE1	diurnal
cce_4173	*cce_4173*	two-component response regulator	diurnal
cce_4183	*cce_4183*	two-component response regulator	diurnal
cce_4190	*cce_4190*	CheA-like two-component hybrid sensor and	insignificant

		regulator	
cce_4195	*cheY*	two-component response regulator receiver domain protein; CheY	circadian
cce_4196	*cce_4196*	two-component response regulator	circadian
cce_4288	*cce_4288*	putative two-component response regulator	insignificant
cce_4289	*cce_4289*	two-component hybrid sensor and regulator	diurnal
cce_4317	*cce_4317*	CheA signal transduction histidine kinases	acyclic
cce_4396	*cce_4396*	transcriptional regulator, ArsR family	diurnal
cce_4405	*cce_4405*	two-component hybrid sensor and regulator	acyclic
cce_4416	*cce_4416*	two-component response regulator	insignificant
cce_4508	*cce_4508*	protein containing a lambda repressor-like, DNA-binding domain	insignificant
cce_4542	*cce_4542*	two-component response regulator	circadian
cce_4543	*cce_4543*	two-component response regulator	acyclic
cce_4578	*cce_4578*	two-component response regulator	diurnal
cce_4579	*cce_4579*	two-component hybrid sensor and regulator	diurnal
cce_4659			
cce_4660	*cce_4660*	protein containing a WD-repeat, a PAS region and a response regulator receiver domain	acyclic
cce_4701	*gst3*	glutathione S-transferase; Gst3	acyclic
cce_4714	*cce_4714*	two-component response regulator, NarL subfamily	circadian
cce_4716	*kaiC2*	circadian clock protein; KaiC2	acyclic
cce_4720	*cce_4720*	transcriptional Regulator, LuxR family	acyclic
cce_4751	*cce_4751*	two-component hybrid sensor and regulator	diurnal
cce_4854	*cce_4854*	transcriptional regulator, ArsR family	
cce_4859	*fur4*	ferric uptake regulation protein; Fur4	insignificant
cce_5029	*cce_5029*	putative sigma factor	acyclic
cce_5088	*cce_5088*	transcriptional Regulator, AraC family	diurnal
cce_5156	*cce_5156*	putative transcriptional regulator	insignificant
cce_5272	*cce_5272*	transcriptional regulator, TetR family	acyclic

Table S3. Comparison of model performance using different clustering methods for definition of target co-expressed clusters.

Distance method	Agglomeration method	TFs	BNs	Number of Clusters 15	20	25	30	35	40	45	50	Mean	Stde
Euclidean	Complete	X	X	0.618	0.524	0.424	0.481	0.464	0.486	0.388	0.368	0.47	0.0
Euclidean	Complete	X		0.653	0.457	0.434	0.357	0.326	0.335	0.292	0.272	0.39	0.1
Euclidean	Complete		X	0.508	0.501	0.438	0.563	0.516	0.536	0.413	0.367	0.48	0.0
Euclidean	Ward	X	X	0.310	0.396	0.432	0.466	0.447	0.455	0.437	0.471	0.43	0.0
Maximum	Complete	X	X	0.275	0.281	0.278	0.341	0.331	0.356	0.356	0.372	0.32	0.0
Manhattan	Complete	X	X	0.256	0.184	0.344	0.373	0.464	0.550	0.483	0.478	0.39	0.1
Canberra	Complete	X	X	0.530	0.436	0.402	0.401	0.493	0.498	0.501	0.481	0.47	0.0
Mink p=2	Complete	X	X	0.607	0.541	0.463	0.487	0.489	0.478	0.371	0.373	0.48	0.0
Mink p=3	Complete	X	X	0.529	0.629	0.461	0.370	0.343	0.365	0.495	0.491	0.46	0.1
Mink p=4	Complete	X	X	0.445	0.440	0.457	0.499	0.503	0.501	0.481	0.473	0.47	0.0
Euclidean	Mcquitty	X	X	0.714	0.704	0.703	**0.746**	0.727	0.719	0.723	0.625	0.71	0.0
Euclidean	Mcquitty	X		0.361	0.631	0.660	0.620	0.579	0.575	0.590	0.643	0.58	0.0
Euclidean	Mcquitty		X	0.700	0.622	0.635	0.695	0.584	0.585	0.578	0.623	0.62	0.0
Euclidean	Mcquitty	X	Rand				0.691						
1-R (Pear)	Complete	X	X	0.410	0.408	0.392	0.382	0.415	0.362	0.390	0.383	0.39	0.0
1-R (Pear)	Mcquitty	X	X	0.440	0.440	0.438	0.437	0.430	0.440	0.431	0.432	0.44	0.0
1-R (Sp.)	Complete	X	X	0.531	0.404	0.334	0.355	0.397	0.369	0.271	0.335	0.37	0.0

Table S4. Validation data set shows low similarity with training data for patterns of expression.
*Correlation was calculated as the minimum or maximum of the Pearson correlation values of the indicated condition (row) with all other conditions not in the same group (indicated at left).

		Model Training Data	
		Correlation*	
Group		Minimum	Maximum
12 h L/D	D1	-0.67	0.99
	D5	-0.56	0.95
	D9	-0.51	0.91
	L1	-0.67	0.98
	L5	-0.54	0.86
	L9	-0.51	0.96
	DD1	-0.68	0.99
	DD5	-0.50	0.95
	DD9	-0.38	0.91
	LL1	-0.68	0.98
	LL5	-0.57	0.86
	LL9	-0.50	0.96
12 h SD	L1	-0.57	0.87
	L5	-0.53	0.83
	L9	-0.45	0.83
	D1	-0.64	0.81
	D5	-0.38	0.81
	D9	-0.35	0.81
	LL1	-0.42	0.87
	LL5	-0.56	0.83
	LL9	-0.38	0.83
	LD1	-0.51	0.44
	LD5	-0.17	0.55
	LD9	-0.45	0.60
6 h LD	L2	-0.30	0.85
	L4	-0.31	0.85
	D8	-0.34	0.95
	D10	-0.38	0.95
	L14	-0.32	0.93
	L16	-0.34	0.93
	D20	-0.45	0.94
	D22	-0.43	0.94
	Model Validation Data		
Low O_2	1h	-0.14	0.27
	2h	-0.27	0.16
	6h	-0.14	0.32