

# Supporting Information

Ashworth et al. 10.1073/pnas.1300962110

## SI Materials and Methods

**Growth and Sampling of *Thalassiosira pseudonana* Cultures.** Axenic cultures of *T. pseudonana* CCMP1335 (Provasoli-Guillard National Center for Culture of Marine Phytoplankton) were grown in enriched artificial seawater (ESAW) medium (1) modified with lower levels of nitrate (80  $\mu\text{M}$ ), silicic acid (110  $\mu\text{M}$ ), and phosphate (20  $\mu\text{M}$ ) and grown in bioreactors. Before the experiment, the diatoms were acclimated to constant 20 °C temperature and ambient  $\text{CO}_2$  under a 12:12 h dark:light diurnal cycle at 125  $\mu\text{mol photons}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$  of photosynthetically active radiation (PAR) provided by fluorescent lamps. Cultures were considered acclimated when the growth rates of three consecutive growth curves were significantly equal to each other. Chemicals were purchased from Sigma-Aldrich. During the experiments the bioreactors were equilibrated at ambient (400 ppm) or elevated (800 ppm)  $\text{CO}_2$  by bubbling of mixed gasses (2 L/min) and stirring (50 rpm) in a 10-L glass bioreactor system (BioFlow).  $\text{CO}_2$ -scrubbed ambient air and pure  $\text{CO}_2$  gas were mixed and regulated using two mass-flow controller/monitors (G265; Qubit Systems) and solenoid controller. The air supply was monitored using a  $\text{CO}_2$  analyzer (model S155; Qubit Systems) and dispersed into the media by bubble injection from orifices in a stainless steel tube at the bottom of the bioreactor with stirring. These sterile, equilibrated bioreactors (8 L of media) were inoculated with 50,000 cells/mL of acclimated, axenic *T. pseudonana* and grown for up to 1 wk on a 12:12 h dark:light cycle. pH was measured continuously with a Mettler Toledo 405 DPAS SC K851425 pH probe, and calibrated with spectrophotometric measurements, taken every 2 h (2). DIC was measured once daily with an Apollo SciTech's DIC analyzer available using reference materials from A. Dickson. Oxygen production was measured continuously with a Mettler-Toledo Sensor inPro 6830 probe. Cellular fluorescent capacity (CFC) was measured using the selective inhibitor of photosystem II, 3-(3,4-dichlorophenyl)-1,1-dimethylurea (DCMU). Samples (300  $\mu\text{L}$   $n = 3$ ) were dark adapted for 5 min and the initial fluorescence was measured in a fluorescence plate reader (BIOTEK FL800) at 680 nm when excited at 488 nm. Thereafter a solution of DCMU was added to the samples to a final concentration of  $3 \times 10^{-5}$  M and the maximal fluorescence ( $F_m$ ) measured after 30 s. Cellular fluorescent capacity was calculated according to the equation  $\text{CFC} = F_m - F_0/F_m$  (3). Specific growth rates were calculated as the natural log of the growth ratio over time ( $k = \ln(N_t/N_0)/(T_2 - T_1)$ ). Cell culture was stained with SYBR Green (Invitrogen) to test for the presence (and ensure the absence of) bacteria. Physiological measurements and cell counts were done every 2 h and cell harvesting for RNA extraction and measurements every 12 h. Nutrients ( $\text{NO}_3$ ,  $\text{NH}_4$ ,  $\text{PO}_3$ , and Si) were sampled twice daily and measured according to ref. 4, at the University of Washington Chemical Oceanography facility.

**Flow Cytometry.** For the measurement of cell division, cells were stained with SYBR green counted with an Influx flow cytometer (Cytopeia). Samples were run directly in the Influx flow sorter without concentration. The sorter uses two solid-state lasers, with an excitation source tuned to emit at 488 nm and a second tuned at 405 nm. The cells were counted based on their forward scatter and red fluorescence characteristics. The green fluorescence emission was measured using a combination of a 500-nm and 600-nm dichroic long-pass filters and a 530/20-nm BP filter attached to a photomultiplier (R928; Hamamatsu). The red fluorescence emission was measured with a 600-nm dichroic

long-pass filter and a 640-nm long-pass filter with a 700/30-nm filter and a red sensitive photomultiplier (R3829; Hamamatsu). The flow cytometer was calibrated with 1- $\mu\text{m}$  fluorescent microspheres (Polysciences). Dividing cells were determined from their cell size and DNA content.

**RNA Extraction, Labeling, and Hybridization.** Cells ( $\sim 3 \times 10^7$ ) were harvested at the end of the dark period (dawn) and at the end of the light period (dusk) by filtration (30–200 mL) and then flash frozen. Total RNA was extracted using the mirVANA kit from Invitrogen (total RNA extraction protocol). Transcribed mRNA was amplified and labeled using the Agilent Quick-Amp Labeling kit (two color). Custom Agilent oligonucleotide array slides (see below) were hybridized with equal amounts of oppositely labeled sample and internal reference RNA according to Agilent array hybridization protocols, and scanned using an Agilent two-color array scanner. Dye-flipped samples and references were hybridized to correct for biases in dye incorporation. In total, 36 arrays were hybridized.

**Gene-Specific Microarray Design.** Gene-specific oligonucleotide arrays were designed and ordered from Agilent Technologies. At least three probes were designed to match the 3' end of each of the 11,390 Thaps3.0 nuclear gene models (5), as well as 180 chloroplast genes (6). This array platform is available in the GEO database (7) under accession GPL16806.

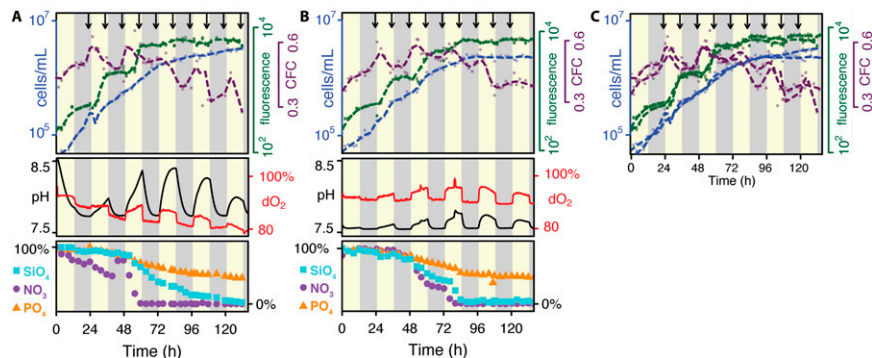
**Microarray Data Processing and Gene Expression Analysis.** Two-color expression data were normalized using the *limma* package in R (8). After normalizing to each uniform internal reference sample, expression ratios were calculated relative to the mean within-experiment expression level for each probe to remove batch effects between experiments. Gene-level correlation between the 400 ppm and 800 ppm expression series was assessed using the Pearson method and compared with a null distribution consisting of  $10 \times 11,390$  randomly resampled gene series pairings. Expression data from all corresponding time series sampling times were analyzed as replicates for the purposes of detecting whole-genome expression patterns that were associated with growth phase transitions. The whole-genome expression data were initially clustered using the K-means method to identify the dominant modes of expression. Subsequently, a two-factor ANOVA (dawn:dusk and exponential:stationary) was performed on the data using the MeV program (9). Exponential samples ( $n = 10$ ) were all of those collected before the light period of day 3 ( $\sim 80$  h). Stationary samples ( $n = 9$ ) were all those collected afterward. This grouping of samples coincided with the decline in specific growth rate of the freely growing cultures to a value below 1/d. Dawn samples ( $n = 10$ ) were collected in the dark after 12 h of constant darkness. Dusk samples ( $n = 9$ ) were collected in the light after 12 h of constant light. The default parameters were used and significance was assessed using an alpha threshold of 0.01 based on 1,000 permutations. Gene functional enrichment within these groups was analyzed using a binomial test with Benjamini–Hochberg correction in Gtools (10) and the DAVID database (11) via the Gaggle framework (12). All microarray data used in this study are available in the GEO database (7) under accession no. GSE45252.

**Transcriptional Regulatory Network Inference.** A model for potential gene regulation by transcription factors on target genes was inferred for each transcription factor in five major families (13). Transcription factors were proposed as candidate regulators for

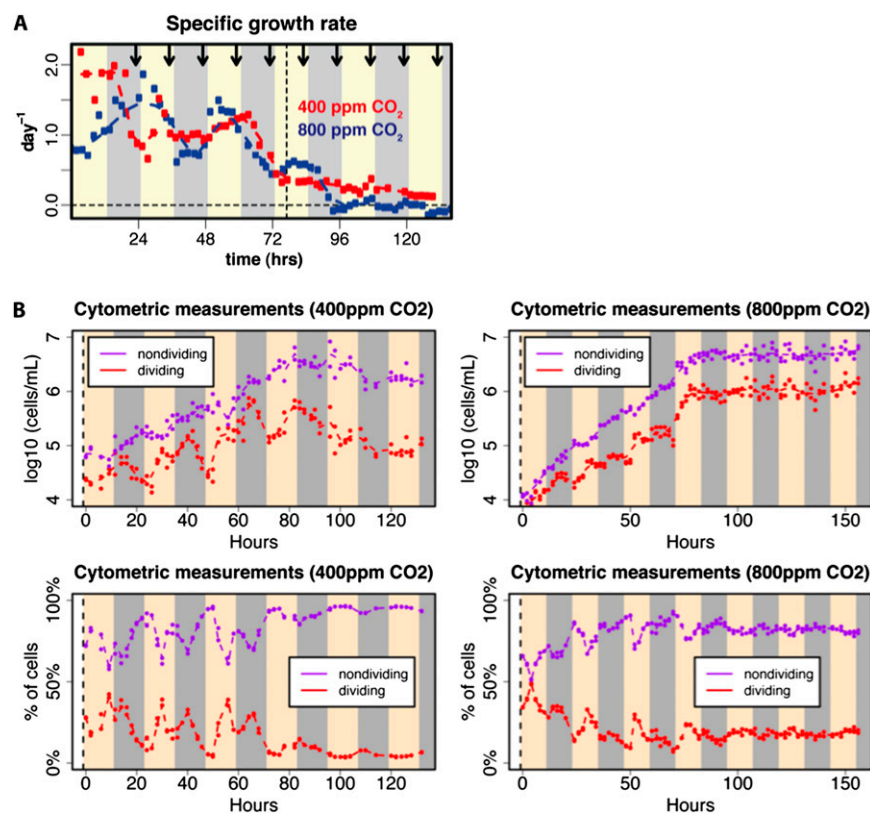
target genes if all of the following were true: (i) the transcription factor (TF) and its target genes were highly correlated in expression changes over multiple published array experiments (two-tailed bootstrap  $P < 0.025$ ); (ii) the upstream region of each target gene (-400 to +50 bp) contains at least one potential DNA binding site that matches a conserved DNA-binding motif for the TF family (motif  $P < 1 \times 10^{-4}$ ); and (iii) the upstream regions of highly correlated target genes were enriched in putative TF-DNA binding sites, compared with all

genes (hypergeometric  $P < 0.05$ ). Putative DNA-binding motifs for conserved TF families were assumed from DNA sequence specificity profiles that are experimentally determined for orthologous transcription factor families in plants. These included the experimentally derived specificity profiles that are representative of the HSF (14), Myb (15), bZIP (16), AP2 (17), and E2F (17) families of transcription factors. The program FIMO (18) was used to detect transcription factor-DNA binding sites in gene upstream promoter regions.

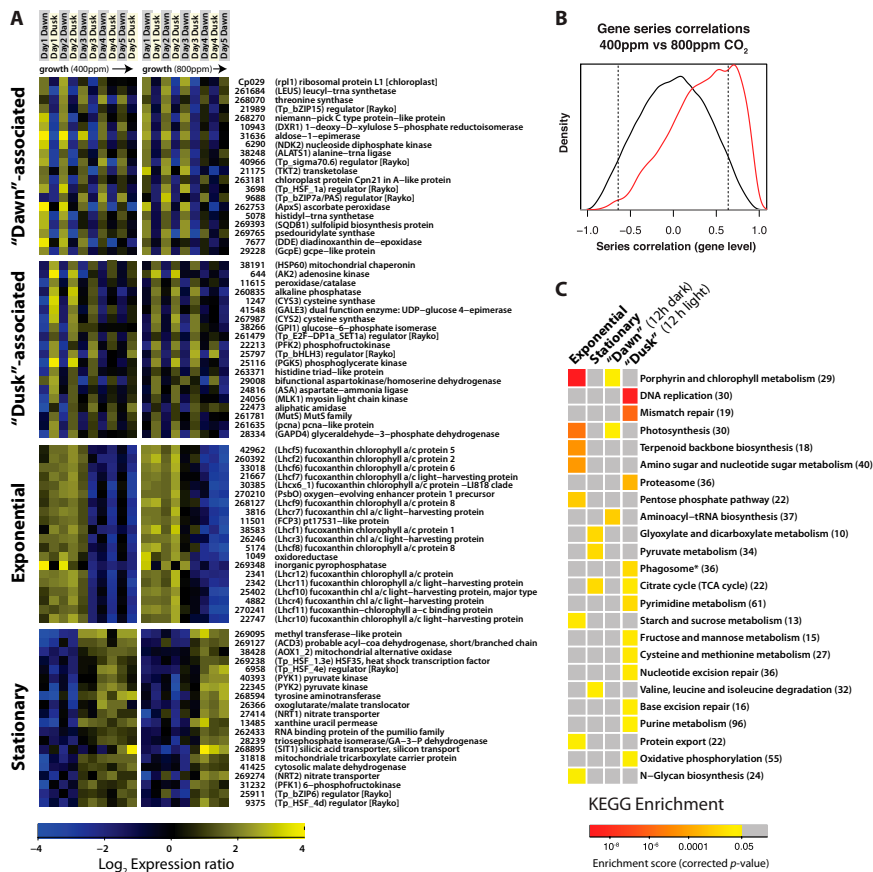
- Berges J, Franklin D, Harrison P (2001) Evolution of an artificial seawater medium: Improvements in enriched seawater, artificial water over the last two decades RID A-4186-2010. *J Phycol* 37:1138–1145.
- Dickson AG, Sabine CL, Christian JR (2007) *Guide to Best Practices for Ocean CO<sub>2</sub> Measurements* (PICES Special Publication, Paris), PICES Special Publication 3, IOCCP Report No. 8.
- Vincent WF (1980) Mechanisms of rapid photosynthetic adaptation in natural phytoplankton communities. II. Changes in photochemical capacity as measured by DCMU-induced chlorophyll fluorescence. *J Phycol* 16:568–577.
- Strickland JDH, Parsons TR (1972) *A Practical Handbook of Seawater Analysis* (Fisheries Research Board of Canada, Toronto).
- Armbrust EV, et al. (2004) The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. *Science* 306(5693):79–86.
- Oudot-Le Secq M-P, et al. (2007) Chloroplast genomes of the diatoms *Phaeodactylum tricoratum* and *Thalassiosira pseudonana*: Comparison with other plastid genomes of the red lineage. *Mol Genet Genomics* 277(4):427–439.
- Edgar R, Domrachev M, Lash AE (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* 30(1):207–210.
- Ritchie ME, et al. (2007) A comparison of background correction methods for two-colour microarrays. *Bioinformatics* 23(20):2700–2707.
- Saeed AI, et al. (2003) TM4: A free, open-source system for microarray data management and analysis. *Biotechniques* 34(2):374–378.
- Perez-Llamas C, Lopez-Bigas N (2011) Gitoools: Analysis and visualisation of genomic data using interactive heat-maps. *PLoS ONE* 6(5):e19541.
- Huang W, Sherman BT, Lempicki RA (2009) Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* 37(1):1–13.
- Bare JC, Shannon PT, Schmid AK, Baliga NS (2007) The Firegoose: Two-way integration of diverse data from different bioinformatics web resources with desktop applications. *BMC Bioinformatics* 8:456.
- Rayko E, Maumus F, Maheswari U, Jabbari K, Bowler C (2010) Transcription factor families inferred from genome sequences of photosynthetic stramenopiles. *New Phytol* 188(1):52–66.
- Storozhenko S, De Pauw P, Van Montagu M, Inzé D, Kushnir S (1998) The heat-shock element is a functional component of the Arabidopsis APX1 gene promoter. *Plant Physiol* 118(3):1005–1014.
- Romero I, et al. (1998) More than 80R2R3-MYB regulatory genes in the genome of Arabidopsis thaliana. *Plant J* 14(3):273–284.
- Martinez-García JF, Moyano E, Alcocer MJ, Martin C (1998) Two bZIP proteins from *Antirrhinum* flowers preferentially bind a hybrid C-box/G-box motif and help to define a new sub-family of bZIP transcription factors. *Plant J* 13(4):489–505.
- Matys V, et al. (2003) TRANSFAC: Transcriptional regulation, from patterns to profiles. *Nucleic Acids Res* 31(1):374–378.
- Grant CE, Bailey TL, Noble WS (2011) FIMO: Scanning for occurrences of a given motif. *Bioinformatics* 27(7):1017–1018.



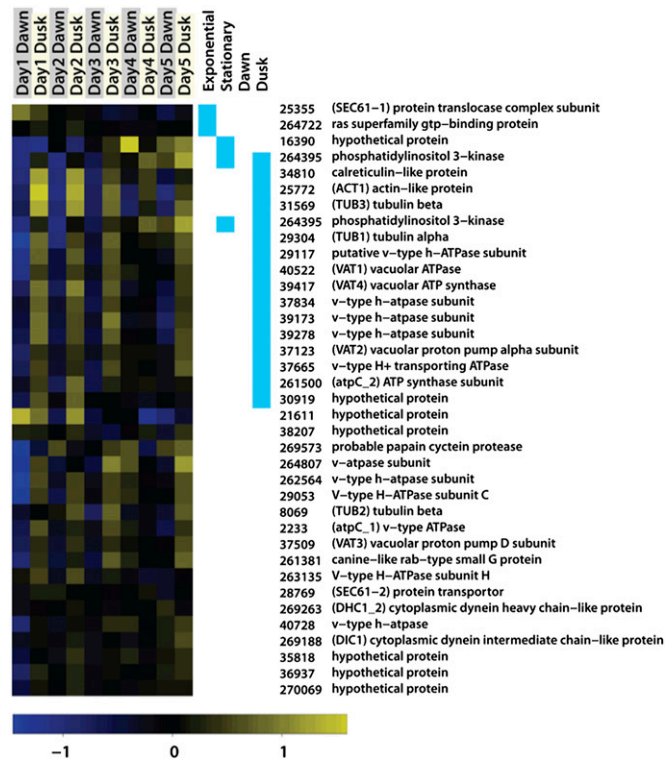
**Fig. S1.** Detailed growth analysis of replicate *T. pseudonana* cultures. Shaded areas indicate dark (gray) or light (yellow) conditions. Upper arrows indicate sampling points for gene expression analysis. (Top) Cell densities are shown in blue, chlorophyll fluorescence (relative units) in green, and cellular fluorescent capacity in purple. (Middle) pH is shown in black and relative O<sub>2</sub> levels are shown in red. (Bottom) Nutrients (NO<sub>3</sub><sup>-</sup>, SiO<sub>4</sub><sup>-4</sup>, and PO<sub>4</sub><sup>-3</sup>) are shown as percentages of the starting concentrations. (A) 400 and (B) 800 ppm CO<sub>2</sub>. (C) Superposition of the cell density, fluorescence, and cellular fluorescent capacity data from A and B.



**Fig. S2.** Specific growth rates and cytometric measurements during replicate growth experiments. (A) Specific growth rates (*SI Materials and Methods*). Arrows indicate sampling times for gene expression analysis. The vertical dashed line indicates the division between the exponential and stationary phases. (B) Cytometric measurements from cell culture during the experiment. Dividing and nondividing cells were estimated by cell size and DNA staining (see *SI Materials and Methods*).

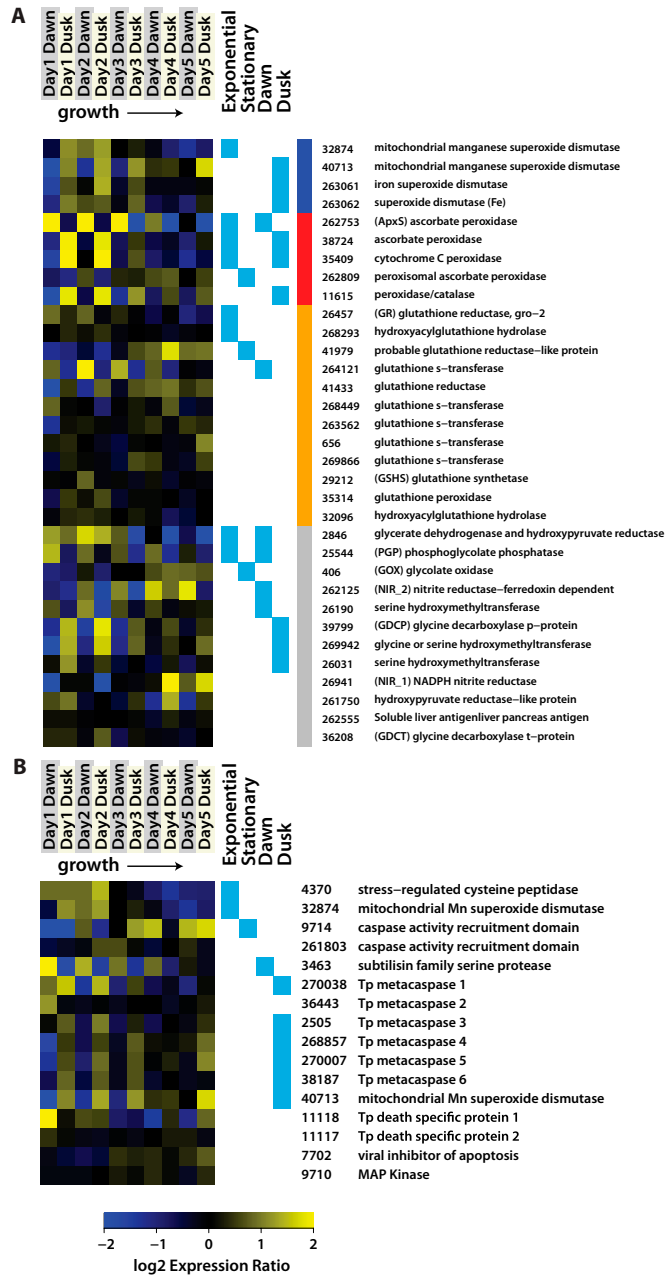


**Fig. S3.** Summary of genome-wide expression data. (A) The top 20 genes for which expression increased most highly during dawn following 12 h of darkness, dusk following 12 h of light, exponential growth, and nutrient-depleted stationary phase. Numbers are Joint Genome Institute (JGI) protein identifiers. (B) Distribution of gene series correlations between 400 ppm and 800 ppm CO<sub>2</sub> replicates. The distribution of gene series correlations between the 400 ppm and 800 ppm CO<sub>2</sub> expression series is shown in red. Null distribution consisting of resampled gene series pairings of the two series is shown in black. Dashed lines indicate alpha levels of 0.05 and 0.95 for the null distribution. (C) The association of functional pathways [Kyoto Encyclopedia of Genes and Genomes (KEGG)] with the four conditions (dawn, dusk, exponential, and stationary), as a function of the enrichment of significantly up-regulated genes. Genes mapping to the phagosome pathway in KEGG (ko04145, marked by an asterisk) include actin, dynein, tubulin, protein transport protein Sec61-1 and -2, and several vacuolar ATPases that are likely involved in cell wall and secretory mechanisms.



**Fig. S4.** Growth phase-associated expression patterns of genes putatively involved in cell morphology, secretion, and cell wall biogenesis. Aqua boxes (*Right*) indicate significant association with the conditional factors according to the analysis of variance. Numbers indicate JGI protein identifiers.





**Fig. S5.** (A) Growth phase expression patterns of genes involved in photorespiration and oxidative stress. *Right* bar is colored according to the following gene groups: blue, superoxide dismutases; red, ascorbate peroxidases; gold, glutathione-related enzymes; and gray, other. *Right* aqua boxes indicate significant association with the conditional factors according to the analysis of variance. (B) Growth phase-associated expression patterns of putative metacaspases and cell death-related genes. *Right* aqua boxes indicate significant association with the conditional factors according to the analysis of variance. Numbers indicate JGI protein identifiers.

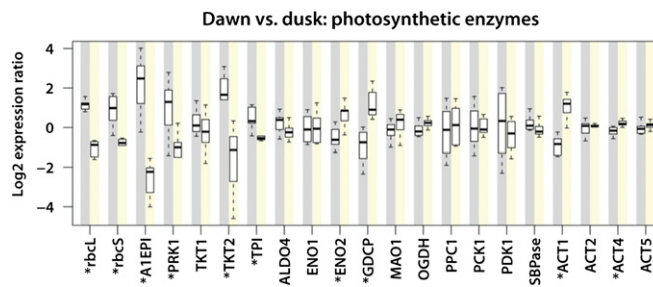


Fig. S6. RuBisCO subunits and several other enzymes related to photosynthesis were differentially expressed between dawn and dusk. Asterisks indicate genes that were significantly differentially expressed ( $t$  test, Bonferroni-corrected  $P$  values  $<0.05$ ).

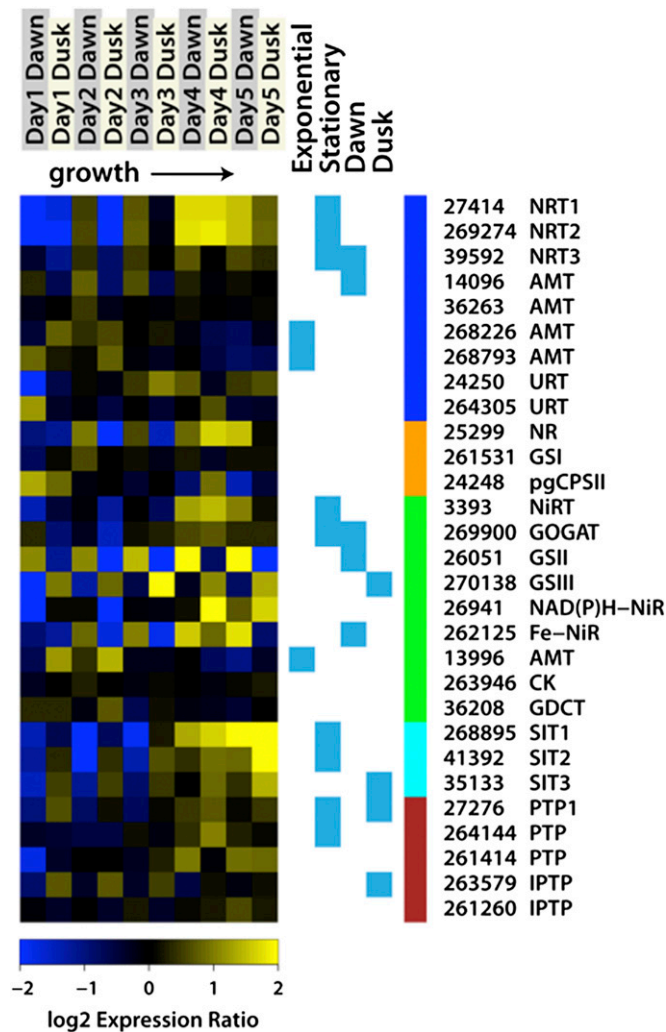


Fig. S7. The expression of genes for nutrient assimilation increases in response to the depletion of their respective nutrients during growth. *Right aqua boxes* indicate significant association with the conditional factors according to the analysis of variance. Colored bar indicates the following gene categories: blue, cell membrane nitrate, urea, and ammonium transporters; orange, cytosolic enzymes; green, plastid-borne genes; cyan, silicate transporters; and brown, phosphate transporters. Numbers indicate JGI protein identifiers.

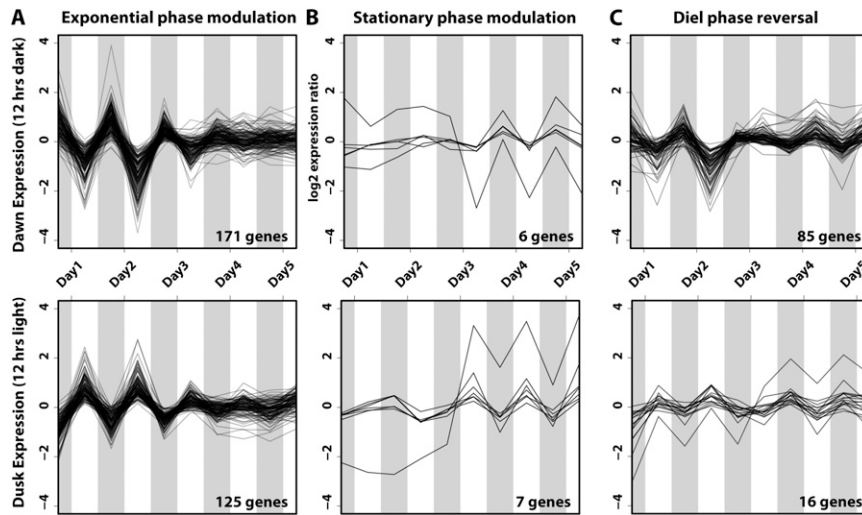


Fig. S8. The expression patterns of several hundred genes were associated with nonlinear combinatorial effects between light:dark and exponential:stationary conditions. (A) Exponential phase modulation occurred for genes that exhibited diel expression changes only during exponential phase. (B) Stationary phase modulation occurred for genes that exhibited diel expression changes only during stationary phase. (C) Diel phase reversal occurred for genes whose diel fluctuations in expression apparently switched from high dawn expression to high dusk expression (or vice versa). Genes with less than twofold changes in diel expression during either relevant phase are excluded from this figure for clarity. See Dataset S1 for individual gene annotations.

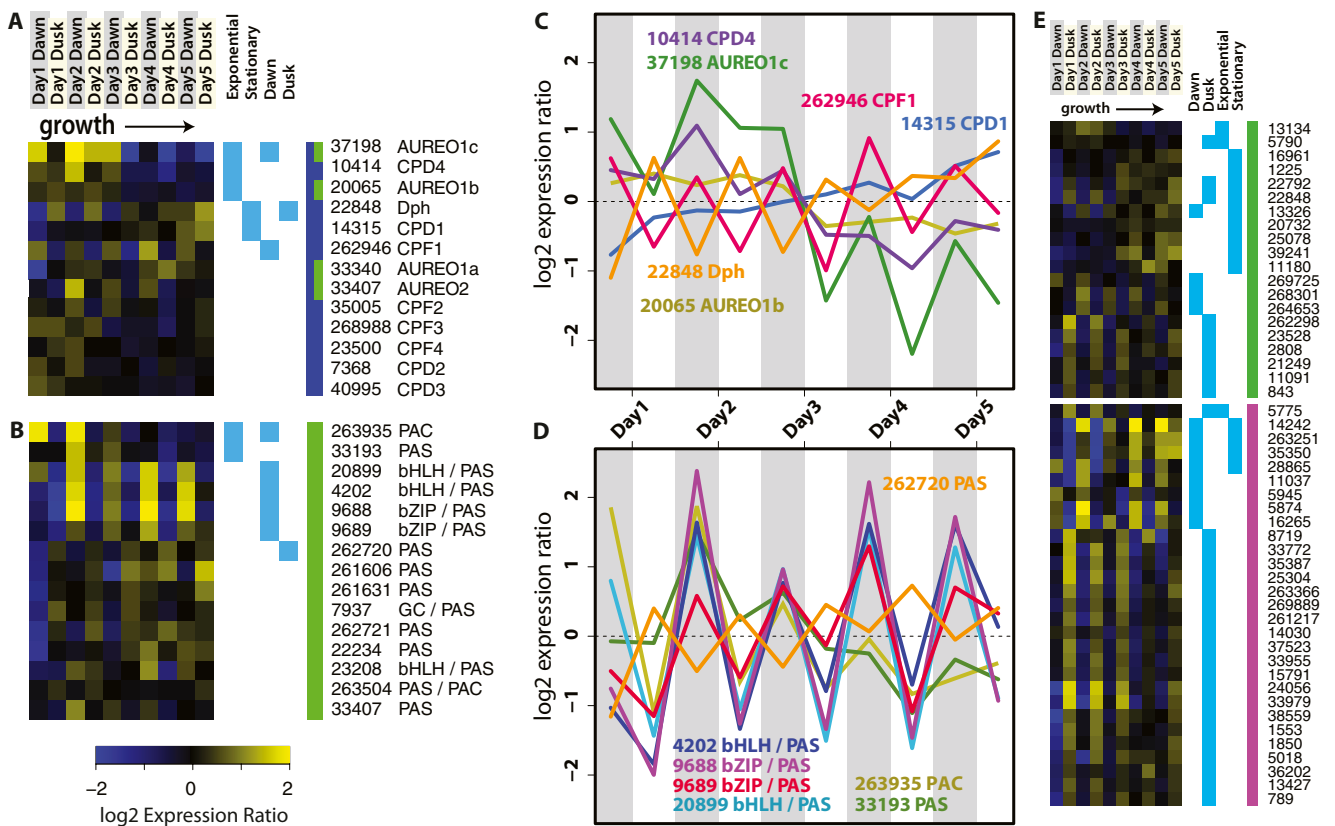
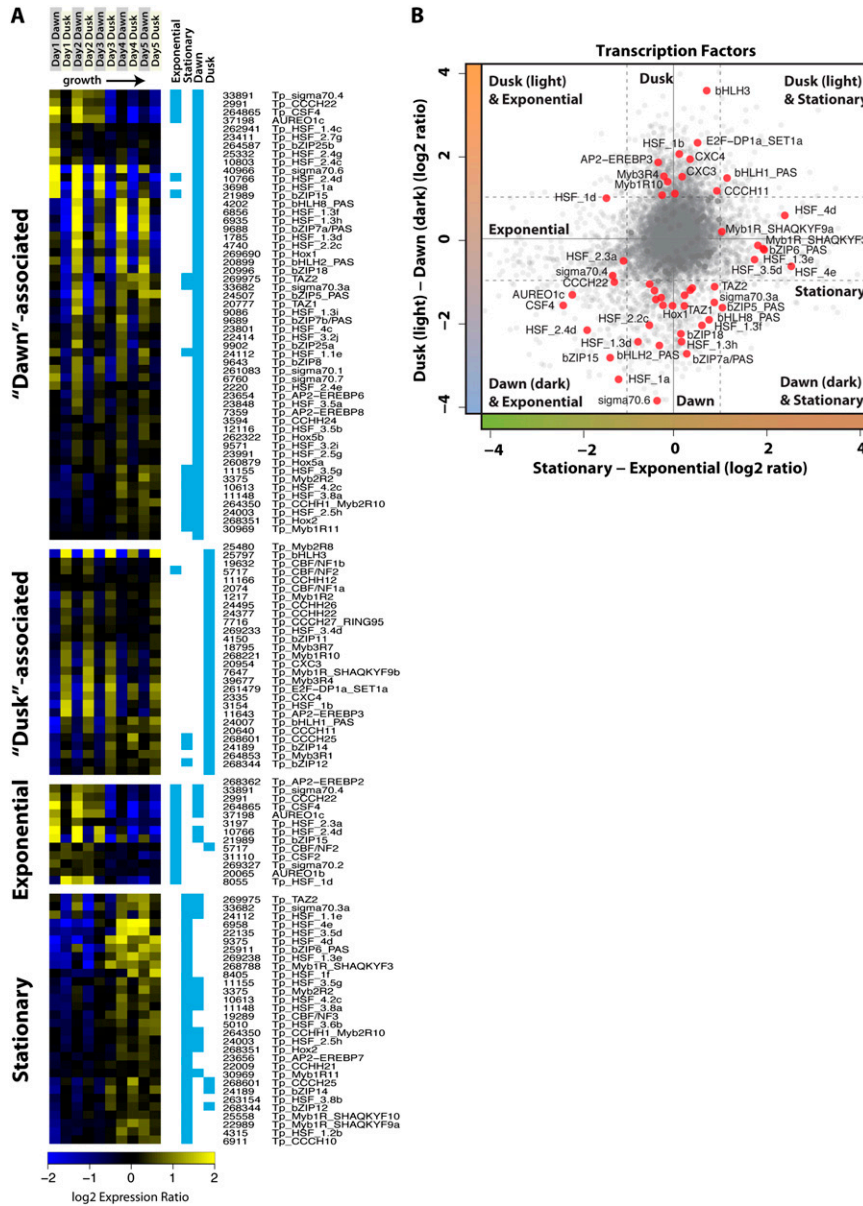
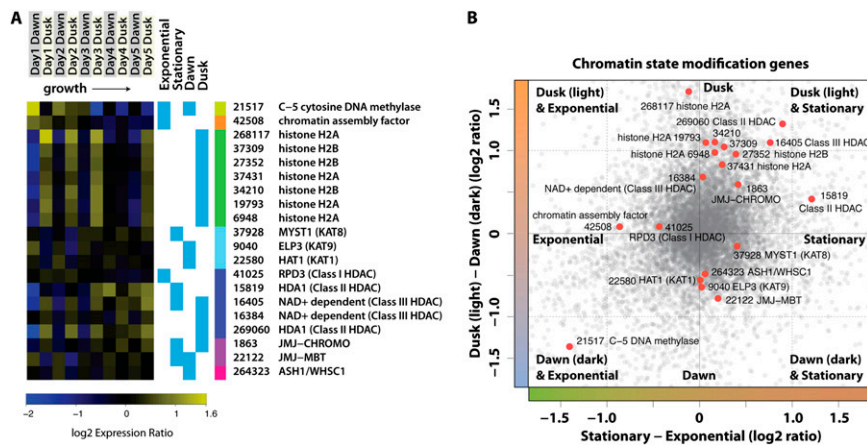


Fig. S9. Expression profiles for putative photoreceptors (A) and signal-sensing genes (B), including those for which expression was not detectably associated with light:dark or exponential:stationary phase transitions. Aqua boxes indicate gene/condition associations by ANOVA at  $P < 0.01$ . Multicolored bar indicates the following classes of genes: putative photoreceptors and photolyases (blue) and PAS/PAC domain-containing genes (green). (C) The expression time series of differentially expressed photoreceptors (Dph, CPF1, Aureo 1B and 1C) and photolyases (CPD1, CPD4). (D) The expression time series for differentially expressed genes containing signal-sensing PAS domains. (E) Putative two-component signaling elements and membrane-bound receptors that were significantly associated with one or more states. Aqua boxes indicate gene/condition associations by ANOVA. Green bar denotes genes containing the following InterPro motifs in conjunction with at least one putative transmembrane domain, as determined by TMHMM v2: protein kinase (IPR000719), response regulator receiver (IPR001789), GAF (IPR003018), GPCR (IPR000337), rhodopsin-like GPCR family (IPR000276), LRR (IPR001611), NB-ARC (IPR002182) and small GTPases (IPR001806). Pink bar indicates putative soluble protein kinases (genes containing IPR00719 motif). Numbers are JGI protein identifiers.

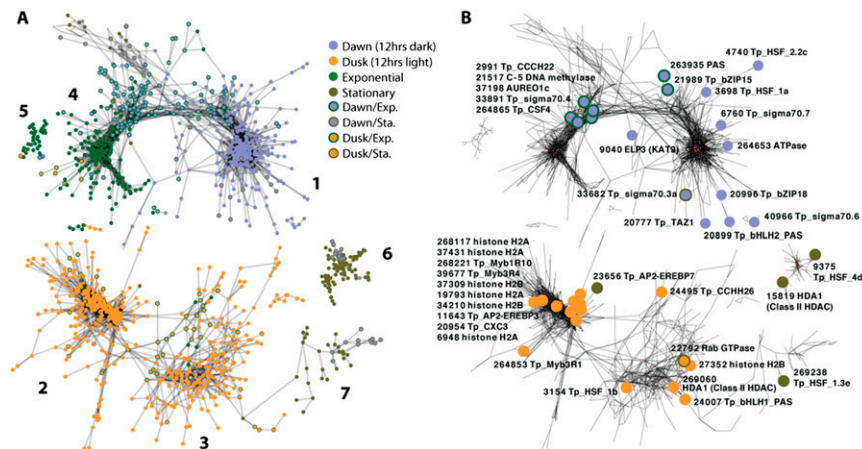




**Fig. S10.** (A) Expression patterns of all transcription factors that were significantly associated with dawn, dusk, exponential, and stationary phases. Aqua boxes indicate gene/condition associations by ANOVA. (B) Transcription factors were differentially expressed along two dimensions: dawn vs. dusk and exponential vs. stationary phases. TFs for which the difference in expression was significant and greater than twofold over at least one transition are highlighted. Numbers indicate JGI protein identifiers.



**Fig. S11.** (A) Genes involved in the regulation of chromatin state are associated with dawn, dusk, exponential or stationary conditions. Colored bars indicate the following gene groups: C5 cytosine DNA methylase (yellow), histones (green), lysine acetyltransferases (light blue), lysine deacetylase (dark blue), lysine demethylases (purple), and lysine methyltransferases (magenta). Aqua boxes in the *Middle* indicate significant association with the conditional factors according to the analysis of variance. (B) Genes with annotated functions for chromatin state regulation were differentially expressed along two dimensions: dawn vs. dusk and exponential vs. stationary phases. Numbers indicate JGI protein identifiers.



**Fig. S12.** Core expression network generated from our data and all other publicly available microarray data. (A) The 1,840 genes that were significantly associated with growth phase transitions during this experiment are also highly correlated over 40 array conditions, including data obtained from refs. 12, 15, 51, and 52 (bootstrap  $P < 0.01$ ). Genes (colored circles) are arranged in a force-weighted network that subdivides the transcriptome into units of conditional coexpression: dawn (1), dusk (2 and 3), exponential (4 and 5), and stationary (6 and 7) phases. The genes and functional enrichments for each of these clusters are listed in [Dataset S2](#). (B) Sensing, signaling, regulatory, and chromatin-modifying genes that are members of these expression clusters. Numbers indicate JGI protein identifiers.

**Table S1. Top 20 genes that were significantly associated and most highly expressed under each of the four phases: (i) dawn, (ii) dusk (iii) exponential, and (iv) stationary**

Gene	Name	Protein	Fold change
Higher expression at dawn			
THAPSDRAFT_11564		Hypothetical protein	23.3
THAPSDRAFT_31636		Aldose-1-epimerase	19.6
THAPSDRAFT_2892		Hypothetical protein	16.5
THAPSDRAFT_40966	Tp_sigma70.6	Regulator [Rayko]	13.0
THAPSDRAFT_22699		Hypothetical protein	11.6
THAPSDRAFT_262753	Apx5	Ascorbate peroxidase	11.4
THAPSDRAFT_269765		Pseudouridylate synthase	11.0
THAPSDRAFT_1734		Hypothetical protein	10.8
THAPSDRAFT_258111		Hypothetical protein	9.0
THAPSDRAFT_6731		Hypothetical protein	8.9
THAPSDRAFT_3215		Hypothetical protein	8.7
THAPSDRAFT_20593		Hypothetical protein	8.6
THAPSDRAFT_3698	Tp_HSF_1a	Regulator [Rayko]	8.6
THAPSDRAFT_25629		Hypothetical protein	8.4
THAPSDRAFT_270308		Hypothetical protein	8.0
THAPSDRAFT_37277		Hypothetical protein	7.7
THAPSDRAFT_22725		Hypothetical protein	7.6
THAPSDRAFT_261684	LEUS	Leucyl-trna synthetase	7.5
THAPSDRAFT_22429		Hypothetical protein	7.4
THAPSDRAFT_6290	NDK2	Nucleoside diphosphate kinase	7.4
Higher expression at dusk			
THAPSDRAFT_6551		Hypothetical protein	9.6
THAPSDRAFT_22860		Hypothetical protein	9.6
THAPSDRAFT_18846		Hypothetical protein	8.9
THAPSDRAFT_25116	PGK5	Phosphoglycerate kinase	7.9
THAPSDRAFT_24649		Hypothetical protein	7.4
THAPSDRAFT_32555		Hypothetical protein	7.3
THAPSDRAFT_7406		Hypothetical protein	7.1
THAPSDRAFT_644	AK2	Adenosine kinase	6.9
THAPSDRAFT_8522		Hypothetical protein	6.5
THAPSDRAFT_34592		Hypothetical protein	6.4
THAPSDRAFT_4536		Hypothetical protein	6.4
THAPSDRAFT_25797	Tp_bHLH3	Regulator [Rayko]	6.4
THAPSDRAFT_262506		Hypothetical protein	6.3
THAPSDRAFT_2376		Hypothetical protein	6.2
THAPSDRAFT_4214		Hypothetical protein	6.2
THAPSDRAFT_5431		Hypothetical protein	6.2
THAPSDRAFT_11569		Hypothetical protein	6.2
THAPSDRAFT_22213	PFK2	Phosphofructokinase	6.1
THAPSDRAFT_10366		Hypothetical protein	6.1
THAPSDRAFT_267987	CYS2	Cysteine synthase	6.0
Higher expression during exponential phase			
THAPSDRAFT_2356		Hypothetical protein	30.9
THAPSDRAFT_3453		Hypothetical protein	18.4
THAPSDRAFT_20603		Hypothetical protein	16.6
THAPSDRAFT_20335		Hypothetical protein	15.3
THAPSDRAFT_7916	FCP_2	Hypothetical protein	14.3
THAPSDRAFT_11557		Hypothetical protein	13.5
THAPSDRAFT_2341	Lhcr12	Fucoxanthin chlorophyll a/c protein	13.3
THAPSDRAFT_1049		Oxidoreductase	13.3
THAPSDRAFT_2673		Hypothetical protein	12.7
THAPSDRAFT_33018	Lhcf6	Fucoxanthin chlorophyll a/c protein 6	12.0
THAPSDRAFT_2342	Lhcr11	Fucoxanthin chlorophyll a/c protein	11.9
THAPSDRAFT_5793		Hypothetical protein	11.3
THAPSDRAFT_38583	Lhcf1	Fucoxanthin chlorophyll a/c protein 1	11.3
THAPSDRAFT_270210	PsbO	Oxygen-evolving enhancer protein 1	11.0
THAPSDRAFT_260392	Lhcf2	Fucoxanthin chlorophyll a/c protein 2	10.9
THAPSDRAFT_22747	Lhcr10	Fucoxanthin chlorophyll a/c protein	10.9
THAPSDRAFT_30385	Lhcx6_1	Fucoxanthin chlorophyll a/c protein	10.8
THAPSDRAFT_270241	Lhcf11	Fucoxanthin-chlorophyll a-c binding prot	10.7
THAPSDRAFT_1989		Hypothetical protein	10.6

**Table S1. Cont.**

Gene	Name	Protein	Fold change
THAPSDRAFT_268127	Lhcf9	Fucoxanthin chlorophyll a/c protein 8	10.4
Higher expression during stationary phase			
THAPSDRAFT_24738		Hypothetical protein	26.5
THAPSDRAFT_269095		Methyl transferase-like protein	16.1
THAPSDRAFT_795		Hypothetical protein	15.4
THAPSDRAFT_20827		Hypothetical protein	13.8
THAPSDRAFT_268594		Tyrosine aminotransferase	11.6
THAPSDRAFT_23160		Hypothetical protein	11.5
THAPSDRAFT_39516		Hypothetical protein	11.4
THAPSDRAFT_4355		Hypothetical protein	11.1
THAPSDRAFT_22731		Hypothetical protein	10.7
THAPSDRAFT_10740		Hypothetical protein	10.7
THAPSDRAFT_8521		Hypothetical protein	10.7
THAPSDRAFT_260934		Hypothetical protein	10.0
THAPSDRAFT_5394		Hypothetical protein	9.8
THAPSDRAFT_5424		Hypothetical protein	9.3
THAPSDRAFT_11569		Hypothetical protein	8.6
THAPSDRAFT_268895	SIT1	Silicic acid transporter, silicon transport	8.5
THAPSDRAFT_4067		Hypothetical protein	8.4
THAPSDRAFT_13485		Xanthine uracil permease	8.4
THAPSDRAFT_22345	PYK2	Pyruvate kinase	8.3
THAPSDRAFT_22671		Hypothetical protein	7.6

**Table S2. DAVID database enrichment categories for condition-associated genes**

Categories	Cluster enrichment score
Genes that differ between dark and light periods 2,846 genes (1,146 up at dawn and 1,700 up at dusk)	
Genes higher at dawn/lower at dusk	
Chloroplast, plastid, translation, rRNA-binding, ribosome, mitochondrion	13.57
Aminoacyl-tRNA ligase and synthetase, amino acid activation, protein biosynthesis	4.41
Transcription factor activity, heat-shock factor, regulation of transcription	1.99
ABC transporter, ATPase activity	1.57
Isoprenoid, terpenoid, carotenoid, pigment biosynthetic processes	1.36
Genes higher at dusk/lower at dawn	
DNA replication, mismatch repair, nucleotide excision repair	3.3
Microtubule, cytoskeleton, motor activity, kinesin	2.63
Proteasome, protein catabolic process, proteolysis	1.78
Genes that differ between exponential and stationary phases 1,747 genes (875 up in exponential phase, 872 up in stationary phase)	
Genes higher during exponential phase/lower during stationary phase	
Photosynthesis, thylakoid, light harvesting, chloroplast, electron transport chain	10.64
Cis-trans isomerase activity	4.94
Polysaccharide metabolic process, chitin binding	3.42
Tetrapyrrole/chlorophyll/pigment metabolic process	3.47
Genes higher during stationary phase/lower during exponential phase	
Glycolysis, gluconeogenesis, citrate cycle	2.67
Glutamate metabolic process, carboxylic acid biosynthetic process	1.51
Transcription factor activity, heat-shock factor, regulation of transcription	1.2
Genes for which the effects of dawn:dusk and exponential:stationary transitions interact 478 genes	
Translation, plastid, rRNA binding, chloroplast, ribosome, mitochondrion	6.76
Aminoacyl-tRNA ligase and synthetase, amino acid activation, protein biosynthesis	5.08
ATP binding, ribonucleotide binding	1.89

Significance ( $-\log(P$  values)) for each gene cluster enrichment is shown in parentheses.

## Other Supporting Information Files

[Dataset S1 \(XLSX\)](#)

[Dataset S2 \(XLSX\)](#)

[Dataset S3 \(XLSX\)](#)

[Dataset S4 \(XLSX\)](#)