High-Resolution Profiling of a Synchronized Diurnal Transcriptome from *Chlamydomonas reinhardtii* Reveals Continuous Cell and Metabolic Differentiation

James Matt Zones,a,b,1 Ian K. Blaby,c,1 Sabeeha S. Merchant,c,d and James G. Umena,a,3

*Division of Biological Sciences, University of California San Diego, La Jolla, California 92095*
*Department of Chemistry and Biochemistry, University of California, Los Angeles, California 90095*

**ORCID IDs:** 0000-0001-8346-7869 (J.M.Z.); 0000-002-1631-3154 (I.K.B.); 0002-2594-509X (S.S.M.); 0000-0003-4094-9045 (J.G.U.)

The green alga *Chlamydomonas reinhardtii* is a useful model organism for investigating diverse biological processes, such as photosynthesis and chloroplast biogenesis, flagella and basal body structure/function, cell growth and division, and many others. We combined a highly synchronous photobioreactor culture system with frequent temporal sampling to characterize genome-wide diurnal gene expression in Chlamydomonas. Over 80% of the measured transcriptome was expressed with strong periodicity, forming 18 major clusters. Genes associated with complex structures and processes, including cell cycle control, flagella and basal bodies, ribosome biogenesis, and energy metabolism, all had distinct signatures of coexpression with strong predictive value for assigning and temporally ordering function. Importantly, the frequent sampling regime allowed us to discern meaningful fine-scale phase differences between and within subgroups of genes and enabled the identification of a transiently expressed cluster of light stress genes. Coexpression was further used both as a data-mining tool to classify and/or validate genes from other data sets related to the cell cycle and to flagella and basal bodies and to assign isoforms of duplicated enzymes to their cognate pathways of central carbon metabolism. Our diurnal coexpression data capture functional relationships established by dozens of prior studies and are a valuable new resource for investigating a variety of biological processes in Chlamydomonas and other eukaryotes.

**INTRODUCTION**

Daily rhythms of biological activities are ubiquitous. For photosynthetic organisms, light and dark cycles are the primary drivers of both metabolism and endogenous circadian clocks that allow organisms to anticipate and adapt to alternating light and dark intervals. Diurnal regulation in plants and algae occurs at multiple levels, including transcription, translation, and posttranscription (Thines and Harmon, 2011; Kinmonth-Schultz et al., 2013; Reddy and Rey, 2014), and a large component appears to be transcriptional as evidenced by transcriptome studies. For example, in the model plant *Arabidopsis thaliana*, 30 to 50% of genes show cyclic diurnal expression patterns under specific conditions of constant temperature (Covington et al., 2008; Michael et al., 2008). Studies on whole plants may be limited by tissue heterogeneity that precludes analyses of cell-type-specific processes unless they are synchronous across most or all tissues (Endo et al., 2014). Unicellular algae offer a complementary approach to understanding diurnal transcriptome regulation as samples contain material from only a single cell type, and mitotic cycles are often amenable to diurnally induced synchronization (Noordally and Millar, 2015). In cyanobacteria, diatoms, and the Prasinophyte alga *Ostreococcus tauri*, up to 80% of genes are periodically expressed in diurnal conditions (Monnier et al., 2010; Johnson et al., 2011; Ashworth et al., 2013). In the red alga *Cyanidiobaschyzon merolae*, a much smaller fraction of genes (~7%) showed diurnal expression periodicity (Kanesaka et al., 2012). These studies revealed a significant temporal segregation of diverse processes, such as photosynthetic metabolism, cell division, and cell behavior (Noordally and Millar, 2015).

For studies of rhythmic or periodic gene expression patterns, two major factors limit the quality of data. The first is sampling density, which determines the temporal resolution of differential gene expression phasing. The second limiting factor is the degree of culture synchrony that influences the coherence of gene expression patterns related to specific processes such as cell division. While some diurnal nuclear transcriptome data have been generated for Chlamydomonas (Kuc, 2005; Panchy et al., 2014), to date there are no deep-sequencing transcriptome studies of Chlamydomonas or other photosynthetic organisms that have combined high-frequency sampling in cultures in which cell synchrony of the entire population of cells was measured and optimized.

Research using the model green alga *Chlamydomonas reinhardtii* (herein referred to as Chlamydomonas) has advanced the understanding of many biological processes, including flagella...
and basal body structure and function, chloroplast biogenesis and photosynthesis, metabolism, DNA methylation, and circadian rhythms (Harris, 2001). Development of Chlamydomonas as a model organism has been further accelerated by the availability of molecular genetic tools and a sequenced genome (Merchant et al., 2007; Umen and Olson, 2012; Jinkerson and Jonikas, 2015). Despite its long history as a model organism, the majority of the predicted proteins in Chlamydomonas have unknown functions, including many that are conserved in other eukaryotic groups (Merchant et al., 2007; Karpowicz et al., 2011).

As observed in other photosynthetic species, cell growth, cell physiology, and mitotic reproduction are strongly influenced by light and diurnal conditions in Chlamydomonas (Howell and Walker, 1977; Mittag et al., 2005; Matsuo and Ishiura, 2010; Mitchell et al., 2014; Tirumani et al., 2014; Cross and Umen, 2015). One advantage of using Chlamydomonas to investigate diurnal or periodic gene expression is the ability to achieve a high degree of culture growth and division synchrony under phototrophic growth conditions (Lien and Knutson, 1979). Strong diurnal synchrony is a natural outcome of the multiple fission cell cycle that Chlamydomonas uses for mitotic reproduction (Bišová and Zachleder, 2014; Cross and Umen, 2015). In a multiple fission cycle, there is a prolonged G1 phase during which cells can grow up to 10 or 20 times in mass before dividing. Following G1, mother cells undergo n rapid, alternating rounds of DNA synthesis and mitosis (S/M phase) to produce 2n uniform-sized daughter cells. Typically, 8 or 16 daughter cells are formed in rapidly growing cultures. In a 12-h-light/12-h-dark regime, cell growth occurs during the day and division occurs at the light-dark transition or early in the dark phase, after which daughters are released and begin growing again during the subsequent light phase.

Here, we have taken advantage of diurnal synchronization in Chlamydomonas to investigate its transcriptome with 1- or 0.5-h temporal resolution across the day-night cycle. The combination of frequent time-point sampling and a high degree of culture synchrony allowed us not only to determine broad patterns of gene expression but also to characterize in detail the temporal ordering of important biological processes as they relate to gene expression patterning, including cell cycle progression, basal body and flagella assembly, ribosome biogenesis, photosynthetic complex formation, and central carbon metabolism. Expression clustering was also employed as a functional classification tool to make and/or validate predictions about protein localization and function. The data set we produced is a powerful new discovery tool and resource for understanding fine-scale temporal patterning of gene expression that occurs during the Chlamydomonas diurnal cycle and cell cycles. It will also serve as a reference data set for comparative studies of coregulated gene expression in algae and other eukaryotes.

RESULTS

The Majority of the Chlamydomonas Transcriptome Is Differentially Expressed in Synchronous Diurnal Growth Conditions

We optimized growth conditions and culture synchrony in order to identify genes that exhibit periodic diurnal and/or cell cycle-regulated expression patterns. Phototrophic cultures were grown in a 12-h-light/12-h-dark diurnal cycle under turbidostatic control to maintain uniform illumination, biomass density, and nutrient levels during the light phase, during which individual cells grew in mass by ~10-fold. Immediately following the dark transition the cells divided synchronously on average three or four times to produce 8 or 16 daughters. Cultures were independently sampled over two 24-h periods at 1-h intervals beginning at ZT1 (Zeitgeber time = 1 h following the onset of illumination). Additional samples were taken at 30-min intervals between ZT11 and ZT15 during S/M phase, for a total of 28 samples per time course, ending at ZT24 (Figure 1A). Culture synchrony was assessed by comparing cell growth in replicate experiments (Figure 1B) and by monitoring progression through two cell cycle transitions: Commitment (reached after the minimum amount of growth necessary to complete one division cycle) and S/M (Figure 1C). The average cell density during the light phase decreased as the average cell size increased so that uniform biomass density and illumination were maintained throughout the growth phase (Figure 1B).

RNA was prepared from the samples and processed for high-throughput transcriptome sequencing to generate more than three million uniquely mapping reads per sample (Supplemental Table 1). The subset of genes that showed the most variable expression levels over each diurnal cycle (coefficient of variation ≥ 1.2) was used to calculate Pearson correlation coefficients of all pairwise combinations between replicate samples (Figure 1D). Replicates were highly correlated (R ≥ 0.97) and could be subdivided into three major groups based on intersample correlation values: (1) light phase (ZT2-ZT12), (2) light-dark transition (ZT12.5-ZT15), and (3) dark phase (ZT16-ZT1) (Figure 1D).

Of the 17,737 predicted genes in Chlamydomonas (Merchant et al., 2007; Blaby et al., 2014), 15,325 had read abundances above a minimal cutoff of 1.061 RPM (reads per million mapped), which was chosen as an inclusion threshold based on reliability for detecting differential expression (see Methods). Of these, 12,592 showed at least a 2-fold change from their mean value in one or more time points across the diurnal cycle with a false discovery threshold of 0.05 and a maximum abundance >1 RPKM (reads per kilobase per million mapped). These genes, representing >81% of the measurable transcriptome, were grouped into 18 major clusters (c1 to c18) based on similarity of expression patterns identified by the K-means algorithm (Soukas et al., 2000) implemented in the MeV software package (Saeed et al., 2003; see Methods), with each cluster containing 250 to 1071 genes (Figure 2A; Supplemental Data Set 1).

We also identified genes with the least variable expression to find those that might be useful for normalization controls. A total of 171 genes had profiles with <2-fold expression variation between maximum and minimum RPKM (Supplemental Data Set 2) and were further examined in 15 other published RNA-Seq experiments (González-Ballester et al., 2010; Castruita et al., 2011; Fang et al., 2012; Urzica et al., 2012b; Malasarn et al., 2013; Goodenough et al., 2014; Schmollinger et al., 2014) (Mn- and Cd-experimental data; S.S. Merchant, unpublished data). We found 24 genes in our low-variance set with maximum fold expression changes of <2.5 and seven genes with fold expression changes of <2 in all other experiments. These stably expressed genes are predicted to be reliable internal controls for gene expression measurements under diverse conditions.
We compared our clustering method to another algorithm, JTK (Jonckheere-Terpstra-Kendall) cycle, which was designed for detecting rhythmic expression from transcriptome data (Hughes et al., 2010). JTK cycle identified 12,534 rhythmic genes with a stringent false discovery threshold (1E-10), 11,142 of which overlapped with our periodic expression data set (Supplemental Figure 1A). In addition, the phasing (i.e., LAG) assignments from JTK cycle were well correlated with specific diurnal expression clusters, meaning that, overall, both methods identified similar periodic expression patterns (Supplemental Figure 1B). However, we also noted a sizeable fraction of genes with either complex or transient expression profiles that were either missed by JTK cycle or given low-confidence scores despite their high reproducibility between replicates and high expression amplitudes (maximum expression/mean expression) (Supplemental Figure 1C). Conversely, the genes identified by JTK cycle that were not included in our differential diurnal set were almost entirely those that cycle reproducibly, but with low amplitude (Supplemental Figure 1D).

Figure 1. Experimental Design and Culture Metrics.
(A) Clock diagram showing experimental design. Diagram shows ZT time of light (ZT0 to ZT12) and dark (ZT12 to ZT24/0) periods of synchronized Chlamydomonas culture with growth (G1), division (S/M), and resting (G0) stages marked by inner gray arrows. Cartoons show relative cell sizes and two successive divisions by multiple fission. Cultures were sampled hourly with additional samples taken at 0.5-h intervals during cell division (brown lines).
(B) Cell size distributions from replicate cultures taken at 3-h intervals during the light phase.
(C) Plots showing fraction of each replicate culture that had passed Commitment and mitotic index at each time point.
(D) Heat map depicting correlation between replicates for the most variably expressed genes (coefficient of variation ≥ 1.2) during the diurnal cycle.
We therefore used our original 18 coexpression clusters as a primary basis for further evaluation.

Functional Annotation of Expression Clusters

To detect functional specialization within the 18 expression clusters, we tested for enrichment of MapMan annotation/ontology terms (Thimm et al., 2004). Term enrichments found in different clusters included protein synthesis (c2, c4, c5, and c12), photosynthesis (c4, c6, c5, and c7), cell division (c10), DNA synthesis and chromatin structure (c1, c9, and c10), abiotic stress (c2 and c7), and cell motility (c11 to c13) (Figure 2B; Supplemental Data Set 3). Many of these ontologies were pursued further using manually curated sets of genes that provided more accurate and comprehensive bases for analysis of coexpression.

Elucidation of Coordinate Expression Patterns and Phase Relationships between Cell Cycle and Chloroplast Division Genes

We extended previous studies of cell cycle gene expression (Bisová et al., 2005; Fang et al., 2006) by first performing comprehensive annotation for 108 Chlamydomonas genes predicted to be involved in cell cycle regulation, DNA replication, chromosome segregation, and chloroplast division (Supplemental Data Set 4). Entry into S/M phase began just after ZT11 and was largely complete by ZT15 (Figure 1C). Strikingly, nearly every one of the cell cycle-related genes showed coordinated upregulation and high peak amplitude just prior to or during S/M phase, with a strong enrichment of replication genes in cluster c10 (Figures 2B and 3A to 3D; Supplemental Figures 4 and 5 and Supplemental Data Set 4). Although individual mother cells do not cycle through each round of S phase and mitosis in synchrony (Fang et al., 2006), phasing of cell cycle gene expression is informative because the first genes expressed will correspond to the earliest cell cycle events (e.g., S phase), while genes that are expressed later will correspond to subsequent events (e.g., mitosis or chloroplast division).

Forty-six DNA replication-related genes showed coordinated expression, peaking at approximately ZT11 near the beginning of S/M, and this coordinated expression serves as a useful temporal landmark for other cell cycle events (Figures 3A to 3D; Supplemental Figures 4 to 6). Cyclin-dependent kinases CDKA1 and CDKB1 are homologs of universally conserved or green lineage-specific cell cycle regulators, respectively (Robbens et al., 2005; Bisová et al., 2005; De Veylder et al., 2007; Cross and Umen, 2015). In Chlamydomonas, CDKA1 function is required to initiate DNA replication, whereas CDKB1 function is required later for mitotic progression (Tulin and Cross, 2014). CDKA1 expression peaked at the time of DNA replication but its lowest expression during G1 phase never dropped below 20% of its peak value (Figure 3E), consistent with CDKA1 protein being detectable in G1 phase cells (Oldenhof et al., 2004). In contrast, CDKB1 expression...
Figure 3. Coordinate Expression and Phasing of Cell Cycle Genes.
is near zero when cells are not dividing and peaked sharply at ZT12, an hour after CDK1 (Figures 3E; Supplemental Figure 7B). Two putative mitotic/S phase cyclin genes, CYCA1 and CYCB1, were expressed maximally at ZT13, just after CDKB1, while the hybrid A-B cyclin gene CYCA81 was expressed maximally at ZT9, 2 h prior to visible signs of cell division (Figures 3E; Supplemental Figure 7B). These data suggest that CYCA81 acts as an early activator of cell division and/or S phase, while CYCA1 and CYCB1 act later as mitotic cyclins. Other conserved genes with predicted roles during mitosis (WEE1, ESP1, and CKS1) were expressed in a pattern similar to CDKB1 (Figures 3F; Supplemental Figure 7C and Supplemental Data Set 4).

The anaphase promoting complex/cyclosome (APC/C) is an E3 ubiquitin ligase that triggers mitotic exit and also serves to maintain low CDK activity in G1 phase (Morgan, 2007; Chang and Barford, 2014). Genes encoding APC/C subunits were expressed in a broad peak from ZT10 to ZT13 but were also expressed detectably in G1 phase (Figures 3G; Supplemental Figure 9A) as was the gene encoding the G1 phase APC/C activator CDH1. In contrast, the gene for the predicted mitotic activator of APC/C, CDC20, was expressed in a sharp peak around ZT13 with little or no expression outside of S/M phase (Figures 3G; Supplemental Figure 9B). These data match key findings in other species where there is a highly specific mitotic requirement for APC/C and a broader requirement for APC/CDC20 and for APC/CDC1H upon mitotic exit and during G1 phase (Zachariae and Nasmyth, 1999).

Proteins of the retinoblastoma (RB)-related protein complex in Chlamydomonas are present throughout the cell cycle where they are required for cellular size control at Commitment and to regulate cell division number during S/M phase (Umen and Goodenough, 2001; Bisová et al., 2005; Fang et al., 2006; Olson et al., 2010). mRNAs for each subunit of the complex (MAT3 [RBR], E2F1, and DP1) are detectable in all samples with peak expression for all three occurring relatively early in the cell cycle at ZT9, consistent with a role in initiation of S phase and/or earlier events such as Commitment (Figures 3H; Supplemental Figure 7A). Interestingly, a gene encoding an E2F-related protein, E2FR1, which has degenerate DNA binding and dimerization domains (Bisová et al., 2005), is expressed prior to the genes encoding the core RB complex, suggesting that this protein has a function that is distinct from that of the canonical RB complex (Figures 3H; Supplemental Figure 7A).

Besides recapitulation of known or anticipated regulatory patterns, our data provide a means for classifying groups of cell cycle regulators that have been less extensively investigated in Chlamydomonas. Whereas most Chlamydomonas cell cycle genes are present in a single copy (Bisová et al., 2005), the D-cyclin family has four paralogs (encoded by CYCD1-CYCD4) whose origins are at least as old as the split between Chlamydomonas and its multicellular relative Volvox carteri (Prochnik et al., 2010). Each D cyclin has a distinct expression profile with peaks at ZT8 (CYCD4), ZT10-ZT11 (CYCD2), ZT12 (CYCD1), and ZT15 (CYCD3) (Figure 3I; Supplemental Figure 7D). These divergent expression patterns point toward D-cyclin subfunctionalization associated with their expression peak phasing: CYCD4 is a candidate regulator of and/or marker for Commitment; CYCD2 is a candidate S phase activator; CYCD1 and CYCD3 are candidate mitotic activators; and the prolonged expression of CYCD3 and CYCD2 also suggests a postmitotic function for these genes.

Unlike land plants that have many individual chloroplasts per cell, Chlamydomonas and most other unicellular Chlorophyte algae have a single large chloroplast that must be physically partitioned during cell division, a process that occurs before cytokinesis (Goodenough, 1970; Gaffal et al., 1995) (Figure 4). Several genes predicted to encode chloroplast division proteins in Chlamydomonas were previously shown to be expressed during S/M (Wang et al., 2003; Adams et al., 2008; Hu et al., 2008; Miyagishima et al., 2012) but were not sampled at high temporal resolution. We found that known chloroplast division genes peak at ZT11-ZT12, consistent with chloroplast division preceding mitosis and cytokinesis, whose cognate genes are expressed at ZT13 (Figures 3J and 4; Supplemental Figure 10). A number of uncharacterized genes with predicted chloroplast-targeting sequences (364 genes) and/or membership in the GreenCut group encoding conserved plant/algal proteins (55 genes) (Karpowicz et al., 2011) were coexpressed with the known chloroplast division genes in clusters c10 or c11 and are candidates for direct participation in chloroplast division or other chloroplast-related processes that are coordinated with the cell cycle (Supplemental Data Set 4).

c10 membership was significantly overrepresented (z-score = 26) within our curated group of cell cycle-related genes identified primarily through homology searches (Figure 2B; Supplemental Figure 11 and Supplemental Data Sets 3 and 4). We asked whether membership in c10 also correlates with cell cycle function in genes identified in an unbiased genetic screen for temperature-sensitive lethal cell cycle mutants (Tulin and Cross, 2014). We found that genes corresponding to the DIV class of mutants, whose defect
was in mitotic progression, were highly enriched for membership in c10 (z-score = 13) (Supplemental Figures 11 and 12 and Supplemental Data Set 5). In contrast, genes for a second class of mutants called GEX that showed a failure to exit G1 phase were found in several clusters and only showed significant enrichment in c1 (z-score = 3.76), which contains a large number of stress-related and chromatin-related genes (Figure 2B; Supplemental Figures 11 and 12 and Supplemental Data Sets 3 and 5). We predict from these data that DIV genes will likely have functions directly associated with cell cycle progression (e.g., DNA replication, mitosis, and cytokinesis), whereas GEX genes represent a more pleiotropic set of functions than DIV genes, and their further study may reveal new connections between the cell cycle and stress pathways and/or chromatin dynamics.

Expression of Genes Encoding Flagella and Basal Body Proteins Is Coordinated with the Cell Division Cycle

The flagella of Chlamydomonas are motility and sensory organelles that are homologous in structure and function to animal cilia and to the cilia or flagella found in other eukaryotic taxa (Carvalho-Santos et al., 2011; Ostrowski et al., 2011). During the cell cycle, flagella/cilia are resorbed or severed prior to S phase or mitosis, thereby freeing basal bodies to act as centrioles during mitosis and cytokinesis (Johnson and Porter, 1968; Coss, 1974; Plotnikova et al., 2009; Parker et al., 2010; Kobayashi and Dynlacht, 2011). Flagella are reformed upon mitotic exit (Wood et al., 2012) and also play a role in daughter cell release as a secretory site for hatching enzyme (Kubo et al., 2009). In our experiment, flagella resorption occurred synchronously as cells entered S/M and flagella reformation was scored upon hatching, by which time daughters already had full length flagella (Supplemental Figure 13).

Although a selected set of genes relating to flagella function were found to be upregulated just after cell division genes (Wood et al., 2012), we wanted to conduct a comprehensive analysis of basal body and flagella gene expression during the cell cycle. We first curated a core set of 193 basal body and flagella genes (Basal Body, Axoneme, IFT, and BBsome) that have been identified in previous studies (Keller et al., 2005; Lechtreck et al., 2009; Dutcher, 2009; Shiratsuchi et al., 2011; Bower et al., 2013) (Figures 5A to 5C; Supplemental Data Set 6). Clusters c11 to c14 are significantly enriched for the MapMan ontology term “cell motility” (z-score > 2.5)(Figure 2B; Supplemental Data Set 3) and contain an overrepresentation of known flagella and basal body genes (P < 0.02) (Figures 6A; Supplemental Figure 14). Remarkably, nearly every flagella and basal body gene exhibited a similar expression pattern with an abrupt rise in expression around ZT12 when S/M was ongoing, peak expression around ZT13-ZT14, and tapering expression during the remainder of the dark period (ZT15 to ZT24) (Figures 5A to 5C; Supplemental Data Set 6). This expression pattern matches the expected requirement for peak flagella synthesis in postmitotic daughter cells. Also notable were the expression patterns of several flagella and basal body genes that differed from the majority (Figures 5A to 5C; Supplemental Data Set 6) and correlate with additional functions for the gene products not involving flagella, e.g., in the actin cytoskeleton (IDA5 encoding actin and PRO1 encoding profilin) and in stress responses (HSP70A, DNAJ1, and HSP90A) (Schroda and Vallon, 2009). Therefore, our data suggest that additional basal body/flagella genes that are not part of clusters c11 to c14 (e.g., POC13/FMO11, 

Figure 4. Summary of Cell Cycle Gene Expression.

Graphical representation of expression phasing for genes or gene groups related to cell division and cell cycle control (ZT11 to ZT14) with darkest colored shading indicating peak expression time. The lower cartoon depicts key stages of division and the relative timing of events within a single division cycle. The backward arrow indicates optional iterations of S phase and mitosis. Additional expression data for cell cycle genes are in Supplemental Figures 4 to 11 and Supplemental Data Set 4.
Figure 5. Flagella and Basal Body Gene Expression.

(A) to (C) Heat maps depict relative expression levels for flagella and basal body genes that are divided into functional subgroups. The maximum expression level for each gene is set to 1. Cell cycle stages, diurnal cycle time, and light and dark phases are indicated above each heat map. Individual gene names are on the right. Cluster membership of genes is shown to the left (gray bars); brown stars indicate nondifferentially expressed genes.

(A) Flagella axoneme genes and subgroups.

(B) Core basal body genes, POC genes, and BUG genes (described in the main text).

(C) Intraflagellar transport (IFT) and Bardet-Biedl syndrome protein complex (BBSome) genes. IFT genes are further divided into Complex A subunits, Complex B subunits, anterograde motor subunits, and retrograde motor subunits.
POC17/PHYB1, and CCT3) (Figure 5; Supplemental Data Set 6) may have additional functions outside of these organelles and that coexpression is a useful discriminatory filter for identifying such genes.

Among the basal body genes (Figure 5B; Supplemental Data Set 6), we noted significant differences in expression patterns and cluster membership among three subgroups: core basal body (those with validated functions), POC (proteome of centrioles), and BUG (basal body upregulated after deflagellation); the latter two groups were identified in a basal body proteomic study, with POC genes validated by homology with centriolar proteins of other species and BUG genes validated by upregulated expression after deflagellation (Keller et al., 2005). We observed that core basal body and axonemal gene groups had distinct distributions among c11 to c14 (P = 6.2 E-09, Fisher’s exact test), with core basal body genes predominantly found in c11 (peak expression at ZT12) and axonemal genes predominantly found in c12 or c13 (peak expression at ZT13 and ZT14) (Figures 5A and 5B; Supplemental Figures 14D and 14E). POC gene cluster membership was not detectably different from core basal body gene cluster membership (P = 0.76, Fisher’s exact test), whereas BUG gene cluster membership was different from both axonemal and core basal body genes, but more similar to the axonemal cluster distribution (P = 0.026, Fisher’s exact test) than to the basal body cluster distribution (P = 8.2 E-05, Fisher’s exact test) (Supplemental Figures 14D and 14E). These data are likely to reflect functional differences and temporal ordering between the core basal body and POC genes that are required during S/M for basal body assembly, replication, and mitosis, and the BUG genes that we predict will encode basal body proteins related to postmitotic nucleation and assembly of flagella (e.g., transition zone proteins). Indeed, many of the BUG genes are also present in the flagella proteome, whereas no POC genes are in this subset (Supplemental Data Set 7). Although the specific functions of most POC genes are unknown, the POC3/CEP290 gene product localizes mainly to the transition zone and is required for flagella assembly, and therefore does not match our prediction for involvement in basal body assembly or replication (Craigie et al., 2010).

Three additional large data sets of candidate flagella/basal body protein coding genes were investigated based on their synchronous expression profiles: genes encoding proteins of the flagella proteome (FAPs) (Pazour et al., 2005), CiliaCut genes that are conserved in ciliated species but missing from species without cilia or flagella (Merchant et al., 2007), and an RNA-Seq-based transcriptome study of genes upregulated after deflagellation (Albee et al., 2013). We asked whether our coexpression data would discriminate between genes encoding proteins with core functions in flagella or basal bodies versus those with additional functions, or in the case of the deflagellation experiment, between genes whose expression is triggered by the stress of deflagellation and genes that are required for flagella biogenesis. First, we found that the uncharacterized members for all three large data sets described above (FAP, CiliaCut, and deflagellation upregulated) are significantly enriched for membership in c11 to c14 (Figure 6A; Supplemental Figure 14C) and predict that the c11 to c14 subset of genes from these data sets has functions directly related to flagella biogenesis.

We extended and validated this observation by examining comembership of genes in the three data sets (FAP, CiliaCut, and deflagellation upregulated) with the assumption that genes found in two or three of the groups are more likely tied to flagellum/basal body function than genes found in only one of the groups. Indeed, we found that members of c11 to c14 were highly overrepresented in the set of genes that belong to more than one data set (P = 1.5 E-27), while genes in other clusters were underrepresented in the overlap set (P = 3.0 E-27) (Figure 6B; Supplemental Figure 15B).

In summary, our data demonstrate robust, coordinated expression timing of basal body and flagella genes that is likely coupled to cell cycle progression and can be used as a powerful filter for identifying and classifying proteins that function in these organelles.

**Ribosomal Protein Genes Are Expressed in Compartment-Specific Temporal Patterns**

Green organisms must coordinate the activity of three separate protein translation systems located in the cytosol, chloroplast, and mitochondria respectively. rRNAs are encoded by the nucleus and respective organellar genomes, but ribosomal protein genes (RPGs) for all three compartments are nucleus-encoded. In principle, RPG expression for the three compartments could be globally coordinated and directly reflects patterns of diurnal cell growth, but instead we found that RPGs were expressed in distinct compartment-specific patterns (Figure 7; Supplemental Data Set 8). Cytosolic RPGs were expressed at high levels throughout the diurnal cycle used in this experiment and as a group constitute between 15 and 45% of all unique protein coding gene transcripts at any given time (Supplemental Figure 16). While cytosolic RPGs almost never dropped below half their peak expression values, they showed a consistent increase during the dark period with peak expression at around ZT15 when cells exited S/M and new daughters were hatching (Figures 7A and 7D). Notably, this was not a time when cells were actively growing. Mitochondrial RPGs showed an almost mirror-image pattern compared with cytosolic RPGs with highest expression in the light and reduced expression in the dark when respiration is most active (Figures 7C and 7D). Most strikingly, chloroplast RPGs showed a very strong periodic expression pattern, with a sharp peak early in the light period at ZT15 followed by a continuous decline through the remainder of the light period and partial recovery during the dark period (Figures 7B and 7D). These expression patterns suggest an unexpectedly strong partitioning of ribosome biogenesis among the three compartments, and they underscore the unanticipated coupling of ribosome biogenesis to organelle-specific requirements rather than cell growth as a whole.

Outside of the general patterns described above, we found that several RPGs that encode plastid-specific ribosomal proteins (i.e.,
PSRPs; Zerges and Hauser, 2009) are expressed out of phase with the majority of chloroplast RPGs. These include PSRP-1, whose expression peaks sharply and transiently at ZT13 but has very little expression at ZT2 during peak expression for most chloroplast RPGs. PSRP-1 is a homolog of the cyanobacterial gene lrt, whose transcript also accumulates after the light-to-dark transition. lrt is thought to stabilize ribosomal subunit association, possibly as a mechanism for stress-regulated translational control (Tan et al., 1994; Samartzidou and Widger, 1998), and the identification of the same pattern in Chlamydomonas suggests a deeply rooted common function for PSRP-1 in the green lineage. RAP38 and RAP41 are related proteins in the GreenCut group of conserved green lineage proteins (Karpowicz et al., 2011) that purify as stoichiometric components of intact Chlamydomonas 70S chloroplast ribosomes (Yamaguchi et al., 2003) and whose genes have nonoverlapping diurnal expression patterns (Figure 7B). Mutants of the Arabidopsis homologs of RAP38 and RAP41 (CSP41a and CSP41b, respectively), have defective rRNA processing and altered polysome profiles, indicating a role for this conserved pair of proteins in rRNA processing and ribosome biogenesis (Beligni and Mayfield, 2008). The noncanonical expression patterns we identified for RAP38 and RAP41 further suggest that they may function differently than core ribosomal proteins and possibly from each other.

In addition to RPGs, we examined expression patterns for other translational machinery and for nucleus-encoded regulators of chloroplast gene expression (Zerges and Hauser, 2009), some of which also showed atypical expression compared with chloroplast RPGs and suggest dynamic changes in chloroplast translation profiles at specific times of the diurnal and/or growth and division cycles (Figure 7B; Supplemental Data Set 8).

In summary, our data support a highly orchestrated pattern of ribosome assembly and compartment-specific translational control. Coexpression profiling further organizes translation-related genes into specific expression subgroups that will provide guidance for future studies of protein biosynthesis.

Plastid and Mitochondrial Protein Complex Subunits Show Phased Expression Patterns

Metabolism is driven by multisubunit protein complexes in chloroplasts and mitochondria, which generate ATP and reductant using either light energy (chloroplasts) or energy from catabolism (mitochondria). Light-driven reactions in the chloroplast are corresponding to those in Figures 5A to 5C, CiliaCut genes, FAP genes, and deflagellation response genes. Enrichment for membership of each gene set within c11 to c14 compared with the entire transcriptome is indicated by asterisks showing significant P values.

(B) Venn diagrams show overlapping membership among indicated categories of flagella-related genes. Cilia Cut, FAP, deflagellation response genes belonging to either c11 to c14 (upper diagram) or all other clusters (lower diagram) were examined. The observed number of genes in a sector is shown inside the sector in bold with the expected membership number and SD based on a null model shown below each number in parentheses. Asterisks indicate sectors whose observed membership deviates significantly from the expected value. Additional data on flagella and basal body genes are in Supplemental Figures 13 to 15 and in Supplemental Data Sets 6 and 7.
Figure 7. Expression Patterns of Genes Encoding Chloroplast-, Mitochondria-, and Cytosol-Targeted Ribosomal Proteins.

(A) to (C) Heat maps depicting relative expression levels of RPGs. Relative expression levels of RPGs and translational regulators targeted to cytosol (A), chloroplast (B), and mitochondria (C). The maximum expression level for each gene is set to 1. Cell cycle stages, diurnal cycle time, and light and dark phases are indicated above each heat map. Gene names are on the right with known targets of chloroplast translational regulators in parentheses next to the corresponding gene name. Brown stars indicate genes without significant differential expression (<2-fold deviation from mean as defined in Methods).

(D) Graph of averaged compartment-specific RPG expression. Average absolute expression levels of large and small subunit RPGs during the diurnal cycle for cytosolic (blue), mitochondrial (orange), and chloroplastic (green) subunits. Additional data for ribosomal protein gene expression and translational regulator gene expression are in Supplemental Figure 16 and Supplemental Data Set 8.
mediated by photosystems I and II (PSI and PSII) and associated light-harvesting complexes (LHCl and LHClII), along with cytochrome b6f and the plastid ATP synthase complex. In mitochondria, the electron transport chain has five major complexes (I to V), including an ATP synthase complex that is encoded by a different set of genes than is the chloroplast ATP synthase. We examined gene expression patterns for the nucleus-encoded subunits for each of these complexes to determine both whether they are coordinately expressed and whether there are peak expression phase differences between them.

Subunits of each photosynthetic complex show broad light-phase expression peaks and are coexpressed in similar patterns, but they also display reproducible intercomplex phasing differences (Figures 8; Supplemental Figure 17 and Supplemental Data Set 9). The earliest expressed genes encode subunits of the ATP synthase and b6f complexes whose mRNA abundances reach their peaks by ZT4-ZT6 and whose basal expression is relatively high compared with genes encoding photosynthetic complexes (Figure 8). Genes encoding subunits of PSI, PSII, and LHCl all show near zero expression in the dark and >1000-fold increases in transcript abundance in the light, with peaks at ZT6-ZT8, representing a 2-h phase delay compared with peak expression of genes encoding cytochrome b6f and ATP synthase subunits. Genes encoding LHClII subunits showed a further phase delay with a peak from ZT8 to ZT10 (Figure 8; Supplemental Figure 17 and Supplemental Data Set 9). The staggered phasing we observed with cytochrome b6f and ATP synthase complexes expressed earliest may enable productive photochemistry without delay upon PSI and PSII assembly, thereby minimizing phototoxic side reactions.

Consistent with previous estimates of protein abundance (Oey et al., 2013; Drop et al., 2014; Natali and Croce, 2015), individual LHC gene mRNA abundances varied by several orders of magnitude compared with the peak expression levels of genes for PSI, PSII, b6f, and ATP synthase complexes, which generally remained within a twofold range between genes (Figure 8; Supplemental Figure 17 and Supplemental Data Set 9). We also observed that genes encoding assembly factors for each photosynthetic complex were either coexpressed with or expressed just before the genes for their target complexes (Supplemental Figure 17 and Supplemental Data Set 9).

When grown photoautotrophically in a diurnal cycle, Chlamydomonas cells accumulate starch during the day through photosynthesis and catabolize it at night via glycolysis and respiration (Gfeller and Gibbs, 1984; Klein, 1987; Thyssen et al., 2001; Ral et al., 2006). The dark-phase respiratory activity associated with starch catabolism is reflected in the expression patterns of genes encoding subunits of mitochondrial electron transport/oxidative phosphorylation complexes I to V. Transcripts for these proteins began accumulating in the mid- to late-light period and peaked in the dark phase (ZT14 to ZT17) (Supplemental Figures 18 to 20 and Supplemental Data Set 10). Thus, respiratory complex genes are expressed out of phase with those of photosynthetic complexes and peak during the dark when respiration is expected to be most active.

**Coordination of Tetrapyrrole Pathway Gene Expression with Light and Dark Phases**

Chlorophyll and heme are the two major classes of tetrapyrrole in photosynthetic cells and both play key roles in energy metabolism. Genes for the common and chlorophyll-specific tetrapyrrole biosynthetic enzymes are coexpressed, with transcript levels increasing rapidly in the light and peaking at ZT6. In contrast, HEM15, which encodes the ferrochelatase enzyme dedicated to heme biosynthesis, shows increasing expression during the dark, highest expression at the end of the dark period, and steadily decreasing expression during the light period when the chlorophyll pathway is most highly expressed (Supplemental Figure 21 and Supplemental Data Set 11). This temporal partitioning between heme and chlorophyll pathway genes may have evolved as a strategy to reduce competition between the two tetrapyrrole biosynthetic pathways and to ensure that chlorophyll is the major tetrapyrrole produced during the light phase (Supplemental Figure 21 and Supplemental Data Set 11).

**Transient Stress Responses Occur at the Dark-to-Light Transition**

The abrupt dark-to-light transition in our experiment enabled us to identify and refine a cluster of 280 genes, mostly derived from c1 and c2, whose expression is transiently induced in a sharp spike at the first light time point, ZT1 (Figure 9A; Supplemental Data Set 12) (see Methods). Functional annotation of this cluster showed enrichment for abiotic stress-related genes (Figure 9B) that include five heat shock protein (HSP) genes and a gene whose Arabidopsis ortholog (At1g04130) encodes a HSP cochaperone (Cre08.g375650) (Schroda and Vallon, 2009). The cluster also included VTC2, which encodes GDP-L-galactose phosphorylase, the enzyme that performs the first committed step in the Smirnoff-Wheeler pathway for ascorbate biosynthesis (Urzica et al., 2012a) (Figure 9A). We hypothesize that many other genes in this cluster are induced by the stress associated with the abrupt transition from darkness into high-light growth conditions.

LHC-like genes in the ELIP, LHCSR, and PSBS families are also associated with light stress responses (Heddad and Adamska, 2002; Hutin et al., 2003; Elrad and Grossman, 2004; Bonente et al., 2008). ELIP1-ELIP5 (encoding homologs of early light inducible proteins) have previously been described in Chlamydomonas (Elrad and Grossman, 2004; Teramoto et al., 2004), and, based on recent genome updates (Blaby et al., 2014), we identified five additional Chlamydomonas ELIP genes, ELIP6 to ELIP10 (Supplemental Figure 22 and Supplemental Data Set 13). ELIP2, ELIP3, ELIP4, ELIP9, and ELIP10 as well as one of two PSBS paralogs, the GreenCut member PSBS2 (Cre01.g016750), are in the light stress cluster. The other PSBS paralog, PSBS1 (Cre01.g016600), showed a similar expression pattern as PSBS2 (Supplemental Data Set 1), but its expression maxima was below the threshold used for inclusion in the cluster. The highly transient light-stress-activated expression of PSBS genes discovered in our study may explain our inclusion in previous experiments and also suggests that, contrary to previous reports (Anwaruzzaman et al., 2004; Bonente et al., 2008; Peers et al., 2009), the Chlamydomonas PSBS proteins may play a role in nonphotochemical quenching (NPQ), the conversion and dissipation of excess excitation energy as heat.

Besides abiotic stress, other functional terms that are enriched in the stress cluster include ABC transporters, protein and amino acid biosynthetic genes, and protein translation (Figure 9B;
Figure 8. Coexpression and Phasing of Nucleus-Encoded Genes for Photosynthetic Complexes.

(A) to (F) Plots of relative expression of photosynthetic genes. Data are averaged for two replicates during light and dark phases (indicated by shading) with expression levels normalized to a maximum of 1 corresponding to the peak expression level of each gene.

(A) LHCII.
Supplemental Data Set 14). This suggests that a concerted metabolic rewiring might occur at the dark-to-light transition when cells must rapidly respond to light stress, while also preparing for a massive upregulation of their photosynthetic capacity in the subsequent light period.

Central Carbon Metabolism Genes Show Pathway-Specific Expression Patterns

Central carbon metabolism in photosynthetic cells must respond to changes in growth conditions such as activity of photosynthetic complexes and external nutrients, and these changes can be reflected by coordinated changes in transcriptional networks for metabolic genes (Wei et al., 2006). The transcript abundance of genes encoding enzymes of the Calvin-Benson-Bassham (CBB) cycle of CO₂ fixation exhibited highly correlated expression patterns along with CP12, encoding a chaperone for GAP3, and the gene for RUBISCO ACTIVASE (Figure 10; Supplemental Data Set 15). Peak expression for nearly all genes encoding enzymes of the CBB cycle occurred between ZT5 and ZT8. An exception from this pattern was RBCS2, one of the two Rubisco small subunit paralogs that was expressed continuously within our diurnal growth conditions. Another exception was TPIC1, the single gene in Chlamydomonas encoding triose phosphate isomerase, which is expressed during the light and dark periods, possibly reflecting its dual roles in the CBB cycle as well as in glycolysis during the dark phase (Supplemental Data Set 15).

Because our cultures were grown photoautotrophically (i.e., without acetate or other organic carbon sources), the patterns we observed for dark metabolism gene expression reflect a state where organic carbon is derived solely from endogenous pools such as starch or storage lipids. Genes encoding enzymes required for the tricarboxylic acid (TCA) cycle and acetate assimilation had less coherent expression patterns than those for the CBB cycle but were still strongly biased toward peak expression during the dark period (ZT12 to ZT24) when respiration dominates energy metabolism (Figure 10; Supplemental Data Set 15). Transcripts for starch biosynthesis enzymes gradually increased during the light period, whereas the expression of genes encoding starch catabolic enzymes was restricted to the dark period when starch is broken down (Supplemental Data Set 15).

Paralog Expression Clustering Predicts Pathway Assignments for Duplicated Metabolic Enzymes

Several enzymes of central carbon metabolism in Chlamydomonas are encoded as multigene families in which individual isoforms may be targeted to different subcellular compartments and are associated with pathways operating at different times during the diurnal cycle (Supplemental Data Set 15). Experimental localization of biochemical activities has been described for some enzymes (reviewed in Johnson and Alric, 2013), but the pathways within which many enzyme isoforms act in Chlamydomonas remain unknown. This uncertainty impacts the accuracy of metabolic models that try to simulate metabolism in a compartmentalized cell (Dal'Molin et al., 2011).

We used expression profiles of “signature” genes that are dedicated to specific pathways (e.g., sedoheptulose-1,7-bisphosphatase [SBP1] to the CBB cycle and isocitrate lyase [ICL1] to the glyoxylate cycle) as a template for correlation measurements that helped assign individual paralogs of duplicated genes to their cognate pathways (Supplemental Data Set 15). For example, citrate synthase activity is required for both the TCA cycle in mitochondria and for the glyoxylate cycle, whose location is presumably in microbodies (the algal equivalent of a plant peroxisome) (Hayashi et al., 2015). The expression profiles of two citrate synthase paralogs encoded by CIS1 and CIS2 were compared with those of the signature TCA cycle genes (SDH3 and SDH4) and to signature glyoxylate cycle genes (MAS1 and ICL1) using Pearson’s correlation. CIS2 expression was strongly correlated with the glyoxylate cycle genes (0.91) and more weakly with TCA genes (0.54), whereas CIS1 expression was poorly correlated with glyoxylate genes (−0.33) and moderately correlated with TCA genes (0.60) (Supplemental Data Set 15). The expression profile-based assignment of CIS2 to the glyoxylate cycle and CIS1 to the TCA cycle was supported and validated by data from Arabidopsis, in which the most similar homologs of CIS2 (CSY4 and CSY5) form a clade of peroxisome-targeted citrate synthases and the most similar homologs of CIS1 (CSY1 to CSY3) form a clade of mitochondrial-targeted citrate synthases (Pracharoenwattana et al., 2005). By applying a similar approach and an expression correlation threshold of ≥0.6, we propose pathway assignments for a total of 23 enzyme isoforms (Supplemental Data Set 15).

Transcription Factor Gene Expression Patterns

We examined the expression profiles of 230 predicted and/or functionally characterized DNA binding transcription factors (TFs) (Pérez-Rodríguez et al., 2010; Jin et al., 2014) in our study and from previously published transcriptome experiments (González-Ballester et al., 2010; Castruita et al., 2011; Fang et al., 2012; Urzica et al., 2012b; Malasam et al., 2013; Goodenough et al., 2014; Schmollinger et al., 2014) (Mn- and Cd- experimental data;
S.S. Merchant, unpublished data; Supplemental Data Set 16). We compared levels of induction, expression amplitudes, and absolute expression maxima as a way of categorizing the regulatory patterns of Chlamydomonas TFs. The mRNA abundances of 219 TFs were $\geq 1$ RPKM in at least one time point of the time course, and of these, 191 showed significant differential expression patterns (Supplemental Figure 23 and Supplemental Data Set 16).

Eighty-six differentially expressed TFs were not significantly upregulated in other experiments/conditions and are therefore possible candidates for controlling diurnal and/or cell cycle related gene expression. In contrast, 105 TFs showed differential expression in our experiment and significant upregulation in at least one other experiment, with 28 of these showing greater upregulation in other data sets compared with their peak amplitude in our data set (max_other/max_diurnal > 2) (Supplemental Data Set 16). Ten TFs were upregulated following deflagellation (Albee et al., 2013), with three of these also members of the core flagella clusters (c11 to c14) and, therefore, candidates for controlling expression of

Figure 9. Transient Stress Response at Dark-Light Transition.

(A) Transcript abundance for known stress response or ROS response genes (red) and GreenCut genes (green). The mean of all 290 light stress gene expression profiles comprising the light stress cluster is plotted in black. Light-gray shading indicates the dark period. Complete protein names encoded by each gene are in Supplemental Data Set 12.

(B) Functional characterization of the light stress cluster. MapMan ontology distributions with percentages of the total are shown in parentheses. Terms highlighted with a star are significantly enriched (P value < 0.05). In total, the light stress cluster includes 13 GreenCut genes, several of which have unknown functions (CGL150, CGL122, CPLD50, and CGL99) (Karpowicz et al., 2011; Heinickel and Grossman, 2013). Additional data on the light stress cluster are in Supplemental Data Set 12.
flagella-associated genes (Supplemental Figure 23 and Supplemental Data Set 16). In contrast, the other seven TFs upregulated in response to deflagellation were also induced in nutrient-limiting conditions, suggesting that they may be part of a more general stress response that occurs after deflagellation.

**Poorly Expressed Genes and Condition-Specific Expression**

The expression estimates for ~2900 predicted genes in our data did not rise above 1 RPKM at any time point (Supplemental Data Set 17). It is likely that the products of these genes are required either at low levels or only under specific conditions that are not part of our diurnal regime (e.g., sexual reproduction and nutrient limitation). Alternatively, some among the 2900 poorly expressed genes may be pseudogenes. We compared expression estimates for these 2900 genes to those from other published Chlamydomonas transcriptomes (Supplemental Data Set 17) and applied a minimum expression estimate of 10 RPKM/FPKM as a threshold for condition-specific expression. In doing so, we identified a total of 460 genes that were significantly expressed in other conditions (Figure 10).

*Figure 10. Expression of Central Carbon Metabolism Genes and Their Diurnal Phasing.*

Metabolic pathways are shown for the CBB cycle (A), acetate metabolism (B), and the TCA cycle (C). Beside each pathway schematic are average expression estimates of the two replicates normalized to peak expression and raw RPKM values for each of the two replicates for each gene of the pathway. Gray shading indicates the dark period. The enzymes catalyzing each step are indicated with primary gene names beside each pathway arrow. Color coding (according to the heat map) and clock icons indicate the time of peak expression. Black arrows and enzyme names in black are used for steps where different isoforms or enzyme subunits are not expressed in the same pattern. Additional data on central carbon metabolism genes including protein names are in Supplemental Data Set 15.
transcriptomes and are therefore strong candidates for being condition-specific genes whose expression is not required for vegetative proliferation (Supplemental Data Set 17).

DISCUSSION

High-Resolution Transcriptomics Reveal Temporal and Functional Relationships among Biological Processes

By combining frequent time-point sampling and high biological synchrony under carefully controlled diurnal growth conditions, we were able to characterize in detail the transcription program of the Chlamydomonas vegetative reproductive cycle with very fine resolution. Using this system, we then investigated diverse biological processes that are diurnally regulated and/or cell cycle regulated (Figure 11). Not surprisingly, we discovered a greater extent of periodic gene expression than in previous studies (Panchy et al., 2014), with ~80% of the detectable transcriptome showing significant differential expression (Figure 2). Just as importantly, we were also able to detect fine-scale phasing patterns that are likely to be meaningful indicators of relative timing among different cellular and metabolic processes or even between sequential events related to the same process. For example, we observed a reproducible ordering of expression of different cell cycle-related genes that matched their predicted order of function (Figure 4). Similarly, we observed statistically significant differences in expression phasing between core basal body genes and flagella structural genes, a finding that is consistent with the temporal ordering of basal body duplication during prophase followed by flagella biosynthesis after cells exit from S/M (Figures 4 and 5) (Wood et al., 2012). Expression of genes for photosynthetic membrane complexes showed phase lags between complexes that directly participate in photochemistry (PSI, PSII, LHCI, and LHCII) versus the βf and ATP synthase complexes whose genes were transcribed earlier than those for PSI and PSII (Figure 8). Moreover, frequent temporal sampling allowed us to discover at least one transient expression cluster containing stress-related genes with a sharp peak at ZT1 that would have been difficult to detect with a sparser sampling regime (Figure 9). Indeed, we observed in this transiently expressed cluster PSBS, whose protein is implicated in light stress signaling and NPQ in land plants, but whose function in Chlamydomonas NPQ has been questioned due to an inability to detect PSBS mRNA under various growth conditions (Bonente et al., 2008, 2012; Peers et al., 2009).

The relative contributions of different external and internal regulatory inputs into the periodic expression profiles we observed remain to be determined. These inputs may include light or dark, cell cycle controls (DNA replication, mitosis, cytokinesis, basal body replication, flagella resorption, and regrowth), and the circadian clock, some of which may be overlapping, as has been observed in Chlamydomonas and other microalgae (Hwang and Herrin, 1994; Serrano et al., 2009; Moulager et al., 2010; Kanesaki et al., 2012; Swirsky Whitney et al., 2012). A genome-wide cDNA microarray-based study in Chlamydomonas found that ~2 to 3% of genes showed detectable circadian cycling under free-running clock conditions, but this is only a small fraction of genes that we found to be periodically expressed under a synchronous diurnal cycle (Kucho et al., 2005). Use of circadian (Matsuo et al., 2008) or cell cycle mutants (Fang et al., 2006; Tulin and Cross, 2014) and alteration of diurnal conditions may help begin to deconvolute these contributions and provide a means of distinguishing how cycling genes are controlled by each input.

Impact of Transcriptome Dynamics on Biological Processes

The mRNA composition of Chlamydomonas cells in our experiment was under constant flux (Figure 2). The widespread finding of co-expression for genes governing related processes suggests that most nuclear genes are under selection for coordinated transcriptional responses at specific times in the cell cycle and/or diurnal growth cycle. We did not assess organellar mRNA levels in our experiment, but these are also known to be periodically expressed and/or diurnally controlled (Salvador et al., 1993; Hwang et al., 1996; Klein, 2008; Idoine et al., 2014).

Flagella are normally made once per diurnal cycle after mitotic exit and basal bodies are replicated during S phase. In these cases, the highly coordinated expression of genes for the constituent proteins of these organelles matches a specific temporal demand. Similarly, most cell division-related proteins are required during S/M and could be detrimental if expressed at other times, so tight temporal coupling between cell cycle mRNA and protein production is also expected, though the magnitude of change we observed for many genes was remarkably high. In cells with well-defined G1, S, G2, and M phases, there are separate transcriptional programs for entry into S phase and M phase (Wittenberg and...
Reed, 2005; Desvoyes et al., 2014), but in Chlamydomonas, the rapid alternation between S and M phases likely necessitates one large burst of transcription for both DNA replication and mitotic genes. Nonetheless, fine-scale expression phasing of genes whose products are predicted to participate in S phase and mitosis are not identical, with replication genes expressed prior to mitotic genes (Figure 4). However, the largely overlapping temporal domains of expression for most cell cycle regulators suggest that if protein abundance is modulated as cells alternate between S and M phases (e.g., cyclin synthesis and degradation), the modulation is likely to be posttranscriptional.

For proteins that are required for photosynthetic growth in the light, such as photosystem proteins, CBB cycle enzymes, and chlorophyll biosynthetic proteins, the production of their mRNAs in the light phase can be rationalized by the demand for increasing chloroplast size and photosynthetic capacity as cells grow. Indeed, our findings match well with direct measurements of photosynthetic metabolites such as chlorophyll and starch from synchronous cultures (Willamme et al., 2015).

On the other hand, RPGs from three different compartments showed maximum expression at different times (Figure 7D). In the case of mitochondrial RPGs, their expression roughly matches the diurnal growth profile of cells. In contrast, chloroplast RPGs show a far more restricted temporal domain of expression, suggesting that most chloroplast ribosomes are produced early in the light phase, perhaps in anticipation of demand for chloroplast translation during growth. However, this initial burst of production is not maintained (Figure 7B). Interestingly, amino acid biosynthetic genes are also upregulated early in the light period as observed by corresponding ontology term enrichments for c4, c5, and c6 (Figure 2B) and the light stress cluster (Figure 9; Supplemental Data Set 14), and this enrichment matches a pattern of free amino acid accumulation early in the light phase (Willamme et al., 2015). The coordinated upregulation of chloroplast protein biosynthetic capacity at a specific time in the diurnal or cell cycle is of potential utility in biotechnology applications that aim to enhance production of photosynthetic metabolites such as chlorophyll and starch from synchronous cultures (Willamme et al., 2015).

The nearly uniform increase in cytosolic RPG expression at the beginning of the dark period is the most difficult to understand because it is out of phase with cell growth. A prior radiolabeling study of cytoplasmic and chloroplast rRNA synthesis and incorporation into ribosomes from synchronous cultures showed steady mass increases of rRNAs and ribosomes in the light phase and higher rates of label incorporation for both compartments during the light phase (Wilson and Chiang, 1977). Thus, RPG mRNA accumulation does not appear to be directly coordinated with rRNA synthesis and ribosome assembly. Ribosomal protein genes can be regulated both transcriptionally and translationally (Mager, 1988; Larson et al., 1991; Perry, 2007) and may also participate in nonribosomal processes including transcriptional and translational control of specific genes (Lindström, 2009). In plants and algae, RPG regulation is less well studied (McIntosh and Bonham-Smith, 2006), especially with respect to diurnal cycles and the cell cycle, though Kucho et al. (2005) did find many chloroplast RPGs in their circadian data set.

A diurnal study of the marine pico-alga O. tauri also found evidence that cytosolic translation machinery might be made in the dark and organelle machinery in the light, though the study only observed a subset of genes and did not discriminate between mitochondrial and chloroplast genes (Monnier et al., 2010). Finding similar dark-phase RPG expression profiles in a distantly related green alga suggests that this expression pattern for cytosolic ribosomes may offer a selective advantage in photosynthetic microalgae, and its underlying bases merit further investigation. In yeast, cytosolic RPG biosynthesis is coupled to G1-S phase cell cycle progression (Jorgensen et al., 2004; Bernstein et al., 2007; Gómez-Herreros et al., 2013), though it remains to be determined if there is any underlying mechanistic similarity between ribosome biosynthesis and cell cycle control in yeast and algae. We also note that at their expression peak, cytosolic RPGs comprise ~45% of all transcripts in Chlamydomonas (Supplemental Figure 16), and their upregulation during the dark phase may reflect the time when RPG expression competes the least with other biosynthetic demands in the cell.

Diurnal metabolism is coupled directly to light conditions, and we found here that genes related to light-regulated metabolic processes were strongly regulated at the level of transcript abundance, including genes for photosystem and light-harvesting complex subunits, chlorophyll biosynthesis, CBB cycle genes, and chloroplast translational machinery whose expression in many cases changed by orders of magnitude between the light and dark phases (Figures 7, 8, and 10; Supplemental Figures 16 to 20). Other areas of metabolism with periodic regulation of expression included respiration/TCA cycle, whose genes were more highly expressed in the dark as expected given their role in starch catabolism. The expression profiles of genes encoding enzymes of other pathways, such as glycolysis/glucogenesis, were less coherent, such as in the case of triosephosphate isomerase (TPIC1), which participates in both anabolic and catabolic processes. In addition, paralogs for central carbon metabolism genes, even those in the same pathway, frequently showed different expression profiles and transcript abundances, suggesting functional partitioning between duplicates or absence of transcriptional control (e.g., RBSC1 and RBSC2; TAL1 and TAL2; and ACS1 to ACS3) (Figure 10; Supplemental Data Set 15). The differential accumulation patterns and relative transcript abundances identified here for enzyme isoforms should help inform metabolic modeling and engineering efforts that require enzyme activity estimates for making accurate predictions, though measurements of protein abundance will be essential to confirm our predictions.

Our findings raise a more general question of how closely the proteome matches that of the transcriptome in synchronous growth conditions and why so many genes are expressed periodically. A global proteomic approach is likely to be informative on this point, but to date has not been done for synchronized diurnal cultures. Very few previous studies on diurnal regulated gene expression in Chlamydomonas investigated the correlation between transcripts and proteins, and most focused on chloroplast proteins and mRNAs that were not considered in our study (Herrin et al., 1986; Eberhard et al., 2002; Lee and Herrin, 2002). Among nuclear genes that have been examined, the highly abundant Rubisco small subunit protein showed elevated abundance in the light and decreased abundance in the dark (Recuenco-Muñoz et al., 2015), similar to the combined mRNA abundance profiles we
measured for its two closely related paralogs, *RBCS1* and *RBCS2* (Figure 10A) (Goldschmidt-Clemont and Rahire, 1986). The mRNA for intraflagellar transport protein IFT27 has a high-amplitude diurnal expression profile with peak abundance just after cell division and very little expression outside this time (Wood et al., 2012) (Figure 5C; Supplemental Data Set 7). Unlike its mRNA, IFT27 protein is present throughout the diurnal cycle but appears to made only once per cell cycle in daughter cells during reflagellation and then becomes slowly diluted as cells grow (Wood et al., 2012). In this case, the protein is more stable than the mRNA, but the periodic mRNA expression still impacts the relative abundance of IFT27 during the diurnal cycle and may even have a global impact on the abundance of other IFT proteins (Qin et al., 2007; Wood et al., 2012).

**Using Transcriptomics of Synchronous Cultures as a Tool for Functional Classification**

Approximately 90% of the genes in *Chlamydomonas* (15,908/17,737) have not been manually annotated with a primary gene symbol, and only ~40% of genes have associated Pfam domains (I.K. Blaby, unpublished data). Our study has revealed that genes for related processes are frequently coexpressed in highly correlated patterns, thus providing a tool that can aid in assigning functions to unknown genes. In each area where we investigated a biological process, our coexpression data not only recapitulated or validated functional information gathered over the past 50+ years (e.g., flagella genes and photosynthetic complexes), but in many cases refined that information and generated new hypotheses about gene function. The utility of our data are exemplified by our analyses of flagella-related genes, for which we not only showed that nearly every known flagella gene is coregulated during the cell cycle, but also classified hundreds of putative flagella and basal body protein coding genes that were identified in large-scale experiments. This will help to narrow down these sets to those likely to be involved directly in flagella and basal body functions. Equally revealing were the few examples of core flagella genes that were outliers and did not display the typical expression patterns of flagella clusters 11 to 14 (Figure 5). In each case where information is available, these outliers correspond to genes with additional functions unrelated to flagella. A second striking example of potentially useful information deriving from coexpression analyses came from chloroplast RPGs and translation factors. These could be grouped into a majority class that were coexpressed with a peak early in the light phase peak and a few outliers whose expression patterns suggest highly orchestrated and specific changes in chloroplast translational regulation over the course of the day–night cycle. Remarkably, our analysis identified *PSPR-1* as a gene whose transient upregulation at the light-dark transition matches that of its distant ortholog in cyanobacteria (Tan et al., 1994; Samartzidou and Widger, 1998), suggesting ancient functional conservation or this ribosomal protein between cyanobacteria and Viridiplantae.

Gene duplication and divergence are well established contributors to evolutionary diversification, and one such type of divergence is in expression pattern between paralogs (Ohno, 1970; Conant and Wolfe, 2008). One of the most challenging problems in metabolic modeling is the assignment of duplicate enzymatic activities to specific pathways and subcellular compartments. Tools designed for predicting protein localization such as Predalgo are valuable but imperfect and sometimes yield results that conflict with experimental data (Tardif et al., 2012). In the absence of data on targeting for over two dozen proteins of central carbon metabolism, coexpression proved to be useful in at least some cases for assigning paralogs to compartments and pathways (Supplemental Data Set 15). It is interesting to note that for some parts of central carbon metabolism, such as the CBB cycle, the genes show highly correlated expression patterns, whereas in other pathways, such as glycolysis and gluconeogenesis, the expression patterns of many genes were discordant. This discordance highlights the limitations of using transcriptome data where in some cases genes for enzymes that participate in the same pathway show uncorrelated or even anticorrelated expression profiles (Figure 10). Such complex profiles could reflect adaptive responses—perhaps resulting from integration with other areas of metabolism that are poorly defined or understood—or could simply reflect lack of selection for a coordinated transcriptional response because enzyme activity regulation is achieved at other levels.

**Compilation of a Reference Data Set for Future Studies of Diurnal and Cell Cycle Regulated Processes**

One potential application for our data set is the construction of transcriptional networks related to diurnal, cell cycle, and/or circadian regulation. Besides uncoupling the inputs corresponding to the above three categories of regulation, a major challenge for such work lies in dealing with highly dynamic data with little or no information available on phase delays between TF expression and target gene expression. The potential lack of correlation between TF mRNA abundance, TF protein abundance, and TF transcriptional activity are additional challenges. Nonetheless, our finding that nearly the entire set of DNA binding TFs in *Chlamydomonas* showed strong and distinctly phased expression patterns suggests that at least some TFs and associated target genes are controlled at the level of transcription. Moreover, we identified a subset of TFs whose transcripts showed dynamic changes in our specific diurnal conditions but not in other transcriptome experiments (Supplemental Figure 23 and Supplemental Data Set 16), making them prime candidates for controlling periodic gene expression. A previous study to identify *cis*-regulatory elements from diurnal transcripts of *Chlamydomonas* proved difficult to compare with our study as culture synchrony in that study was not measured, coregulated gene lists were not published, and the “gold standard” genes used to validate clusters contained some entries that were duplicated between clusters (Panchy et al., 2014). Based on our findings, the identification of *cis*-regulatory elements in *Chlamydomonas* may not be trivial because multiple TFs are expressed in each cluster and each might have its own distinct binding motif. Nonetheless, our data provide an improved resource for attempting to find *cis*-regulatory elements and associate them with cognate TFs.

Our analyses of a high-resolution synchronous transcriptome from *Chlamydomonas* covered several key areas of cell biology and metabolism (Figure 10) but were by no means exhaustive. There are many other cellular processes whose investigation will
be enabled by our data. A range of cell behavior and physiology in Chlamydomonas is under cell cycle, diurnal, or circadian control (Matsuo and Ishiura, 2010). Given the prevalence of transcript cycling (>80% of expressed genes) and coexpression of genes with related functions, it is likely that many processes in Chlamydomonas have some periodic component that can be investigated productively using data from our study.

METHODS

Culture Conditions

Strain CC-5152 was derived from a cross between parental strains 21gr (CC-1690) and 6145c (CC-2895) and used for all experiments. CC-5152 was inoculated from a single colony into a sterilized 400-mL FMT150 photobioreactor (Photon Systems Instruments) containing HS culture media (Davis, 1989). The culture was grown at 28°C in diurnal (12 h light, 12 h dark) conditions with equal fluence of red (630 nm) and blue (450 nm) light (125 µM m⁻² s⁻¹ each) and maintained at a density of OD₆₀₀ = 0.3 ± 0.01 and volume of 400 mL by automated addition of fresh media and removal of excess culture.

Sample Collection and RNA Preparation

Cultures were equilibrated in the photobioreactor for 3 to 4 d prior to sampling. Sampling volumes were adjusted as described below to ensure that equal biomass (pellet wet weight) was present in each sample. Pellet biomasses were each equivalent to ~2 × 10⁶ daughter cells. For light phase samples, 30 mL was taken at each time point and culture volume was restored to 400 mL by automated dilution as the remaining culture grew. For dark phase time points, media were replaced after each sampling to maintain a 400-mL culture volume, and progressively larger volumes were removed at successive time points to compensate for culture dilution. A small sample of culture from each time point was fixed in 0.2% (v/v) glutaraldehyde and 0.005% (v/v) Tween 20 and used to measure cell size and concentration using a Coulter Counter (Beckman Coulter; Multisizer 3). Mitotic progression and passage through commitment were scored as previously described (Fang et al., 2006). Presence/absence of flagella was scored microscopically using the same fixed samples as for scoring mitotic progression. Replicates were performed one month apart from each other.

Collected cells were pelleted by centrifugation (30 s, 3220 RCF, 24°C) in 0.005% (v/v) Tween 20 and resuspended in 0.25 mL RNase-free water and then mixed with 0.25 mL lysis buffer (50 mM Tris-HCl, pH 8.0, 200 mM NaCl, 20 mM EDTA, 2% SDS, and 1 mg/mL Proteinase K) at 70°C. Ten milliliters of this mixture was mixed with 0.25 mL lysis buffer (50 mM Tris-HCl, pH 8.0, 200 mM NaCl, 0.005% (v/v) Tween 20 and resuspended in 0.25 mL RNase-free water and incubated at 70°C for 10 min. This was followed by another purification using AMPure SPRI beads at a ratio of 140:100 beads:DNA volume. Second-strand cDNA synthesis was performed using DNA Polymerase I (Invitrogen), RNase H (Invitrogen), and a nucleotide mix containing dUTP (Roche). The second-strand synthesis was done at 16°C for 60 min. The double-stranded cDNA fragments were purified using AMPure SPRI beads at a ratio of 75:100 beads volume: cDNA volume followed by a second purification using a bead:cDNA ratio of 140:100. cDNA fragments were end repaired, phosphorylated, and A-tailed using a Kapa Biosystems kit followed by ligation to Illumina barcoded sequencing adapters. AmpErase UNG (uracil N-glycosylase; Applied Biosystems) was added to the double-stranded cDNA library fragments and incubated at 37°C for 15 min to cleave and degrade the strand containing dUTP. The single-stranded cDNA was then enriched using 10 cycles of PCR with Illumina TruSeq primers and purified using AMPure SPRI beads at a ratio of 90:100 beads volume:DNA volume to create the final cDNA library. Libraries were quantified using KAPA Biosystem’s next-generation sequencing library qPCR kit using a Roche LightCycler 480 real-time PCR instrument. The libraries were then prepared for sequencing on the Illumina HiSeq sequencing platform using a TruSeq paired-end cluster kit, v3, and Illumina’s cBot instrument to generate clustered flow cells for sequencing. Sequencing of the flow cells was performed on the Illumina HiSeq2000 sequencer using a TruSeq SBS sequencing kit with 200 cycles, v3, following a 2 × 100 indexed run recipe. Primary data are available at the NCBI Gene Expression Omnibus repository under accession number GSE71469.

Sequence Analyses and Identification of Genes with Differential Expression

Quality control of RNA-Seq samples was performed on the raw paired-end reads using Trimmomatic (Bolger et al., 2014) to remove contaminating Illumina adapter sequences and low-quality sequences (average Q20 over a 4-base sliding window <20). Paired reads that were <25 bp at either end were discarded. Reads were aligned to the Chlamydomonas reinhardtii genome v5 assembly with STAR (Dobin et al., 2013) using standard presets except for intron size, which was set between 20 and 3000 bp (–alignIntronMin 20 and –alignIntronMax 3000). Greater than 10 million reads were mapped for each sample (Supplemental Table 1) with uniquely mapping reads accounting for ~90% of total mapped reads in each sample. Uniquely mapping reads were assigned to 17,737 version 5.3.1 primary transcripts using HT-Seq (Anders et al., 2015). One read was added to each transcript in each sample before normalization to allow computation of expression ratios across all samples. This addition had no measurable effect on normalized gene expression estimates. Expression estimates were normalized to library size (uniquely mapping reads per million [RPM]) and pairwise Pearson correlations of normalized samples were computed to evaluate replicability. All replicates were highly correlated (R > 0.973) except for the pair of 6-h dark samples taken at ZT18 (R = 0.938). To determine whether one or both of these replicates was a possible outlier, each was compared with adjacent time points by Pearson correlation, ZT17 and ZT19, with the assumption that high similarity between adjacent samples is an indicator of sample integrity (see Figure 1D showing correlations for all pairwise sample comparisons).
the two 6D samples, 6D_1 was well correlated with both replicate samples at times ZT17 (R > 0.99) and ZT19 (R > 0.976). The 6D_2 sample was less correlated to the ZT17 and ZT19 samples (R = 0.927) and was discarded as a technical outlier. Because a minimum of two replicates per sample is required for DESeq2 differential expression analysis, a replicate of 6D_1 was mapped and quantified in place of the 6D_2 sample. HTSFilter was used to establish a minimum expression threshold of 1.061 average RPM for detection of differential expression (Rau et al., 2013). A total of 2412 genes that did not meet this criterion were removed from consideration.

Differential expression analysis was performed on the remaining 15,325 RPM-normalized genes using DESeq2 (Love et al., 2014) by comparing the replicate average for each gene in each sample with the average normalized expression for the gene across all samples. A total of 13,118 genes were classified as differentially expressed based on having a greater than 2-fold expression difference from their mean expression in at least one time point, with a false discovery rate of <0.05. After differential expression analysis, expression estimates were normalized by the primary transcript length for each locus to calculate RPKM values. A total of 526 differentially expressed genes with a maximum RPKM <1 were removed from further analyses, leaving 12,592 significantly differentially expressed genes, 2179 nondifferentially expressed genes, and 2966 nonexpressed genes.

Coexpression Clustering

A modified implementation of the KMC K-means algorithm implemented in MeV was used to cluster the 12,592 significantly differentially expressed genes (Saeed et al., 2003). First, replicate expression estimates were averaged and then transformed to standard deviations from the mean expression for each gene. The figure of merit algorithm was used to estimate an appropriate number of clusters (Yeung et al., 2001). K-means support using Pearson’s correlation was then used to separate groups of coregulated genes into six initial clusters. Genes that did not cocluster with the main clusters were grouped into smaller satellite clusters. A subsequent round of figure of merit and K-means support was performed on the six main clusters with correlation thresholds varying between 0.7 and 0.96 and the number of iterations varying between 25 and 100, to produce 18 final clusters. Centroids were computed for each of the 18 clusters and genes from remaining satellite clusters were assigned to their closest match among the 18 centroids.

JTK Cycle Analysis

The program JTK cycle (Hughes et al., 2010) was used to identify rhythmic genes as follows. The two replicate data sets were treated as experimental duplicates. The data were filtered to include only the 14,771 genes that met minimum expression criteria (maximum average expression > 1.061 RPM and > 1 RPKM). The temporal spacing between samples was set to 0.5 h and the target period was set to 48 h (effectively equal to 24 h). A false discovery rate cutoff (BH.Q < 1E-10) was chosen to extract a set of 12,534 rhythmically expressed genes that was comparable in size to the differentially expressed gene set identified as described above and used for clustering.

Functional Annotation of Expression Clusters

Functional annotations for Chlamydomonas reinhardtii v5.3.1 predicted proteins were obtained from the MapMan website (http://mapman.gabipd.org) (Thimm et al., 2004) and converted to v5.5 locus IDs (referred to in this study) based on a correspondence table downloaded from Phytozome 10.1 (http://phytozome.jgi.doe.gov). Level 1 and 2 MapMan ontology terms were tabulated for the entire expressed transcriptome of 14,771 genes (>1.081 RPM and >1 RPKM) and used to generate a background distribution model for clusters c1 to c18 along with the 2179 nondifferentially expressed genes that met the expression threshold criteria above. A total of 10,000 data permutations were created by random sampling without replacement and cluster assignment (formally equivalent to generating a hypergeometric distribution). The mean (TermOccurrence_clusterMean) and sd (TermOccurrence_clusterSD) for each term were calculated from this background distribution and used for comparison with actual distribution data. Z-scores (z-score_term,cluster), P values (p-value_term,cluster), and false discovery rates with correction for multiple testing (q-value_term,cluster) were calculated for MapMan terms in each cluster using R functions, as follows:

\[
\text{z-score}_{\text{term,cluster}} = \frac{\text{TermOccurrence}_{\text{cluster}} - \text{TermOccurrence}_{\text{clusterMean}}}{\text{TermOccurrence}_{\text{clusterSD}}}
\]

\[
\text{p-value}_{\text{term,cluster}} = 2 \cdot \text{pnorm}(-\text{abs}(z\text{-score}))
\]

\[
\text{q-value}_{\text{term,cluster}} = \text{p.adjust}\left(\text{vector of all p-values}, \text{method} = "\text{FDR}\"\right)_{\text{cluster}}
\]

A MapMan term was considered to be significantly enriched in a cluster if its maximum z-score was >1.96 (corresponding to a P value < 0.05) and the corresponding false discovery rate was <0.05. Terms associated with fewer than six loci were not considered.

Gene Identifier Conversions

Reads were mapped to genome version 5.3.1 (v5.3.1) available at Phytozome (http://phytozome.jgi.doe.gov), and expression estimates were assigned to v5.3.1 gene loci. Conversion between 5.3.1 IDs, the newest set of IDs from v5.5, and IDs used in previously reported experiments were done using a correspondence table available from Phytozome (Chlamydomonas Transcript Name Conversion Between Releases. Mch12b.txt). Four v5.3.1 models were each split into two new genes in v5.5. For our analyses, we merged the two v5.5 loci back into a single synthetic gene model as follows: v5.3.1 locus, Cre03.g167650, and synthetic v5.5 locus, Cre03.g167622_Cre03.g167644; v5.5.1 locus, Cre14.g613050, and synthetic v5.5 locus, Cre14.g613050_Cre14.g613075; v5.5.1 locus, Cre16.g681700, and synthetic v5.5 locus, Cre16.g681600_Cre16.g681700; v5.3.1 locus, Cre17.g743300, and synthetic v5.5 locus, Cre17.g743288_Cre17.g743307.

For flagella proteome genes (those with more than two mapped peptides) (Pazour et al., 2005), Cilia Cut genes (Merchant et al., 2007), deflagellation upregulated genes (Albee et al., 2013), and transcription factor genes (Pérez-Rodriguez et al., 2010; Jin et al., 2014) (Supplemental Data Sets 7 and 16) that did not have a single v5.3.1 ID, we did a manual BLASTP search against the v5.3.1 proteome to identify a corresponding gene model based on high sequence coverage (>50%) and retention of any signature sets present in the original gene model. For all other transcriptome comparisons, only IDs with 1:1 correspondences with v5.3.1 ID were used.

Tests for Cluster Membership Enrichment

Enrichment tests of cluster membership for annotated cell cycle and flagella genes were performed using a resampling method similar to that used for MapMan Term enrichment (see above). A background distribution of cluster membership for sample sizes equal to the number of cell cycle genes or flagella genes was computed and used to derive expected cluster membership numbers and associated statistics.

Comparisons of cluster membership differences between gene groups were performed using Fisher’s exact test. The 5×2 contingency tables were constructed using observed measurements in individual clusters versus a background distribution based on expected values from
resampling. The flagella cluster group was comprised of genes in clusters c11 to c14, with genes from all other clusters plus unclustered expressed genes combined into a second group for each test.

Comparisons of cluster membership differences between Basal Body, BUG, POC, and Axoneme gene groups were also performed using Fisher’s exact test. A 4 × 5 contingency table was constructed using observed measurements in clusters and unclustered expressed genes. Clusters c11 to c14 were treated as individual groups and membership in all other clusters and unclustered genes were combined into a fifth group.

Enrichment of Cluster c11 to c14 Membership in Data Sets Containing Predicted Flagella Genes

Candidate flagella genes from the flagella proteome (Pazour et al., 2005), CiliaCut (Merchant et al., 2007), and deflagellation response (Albee et al., 2013) were scored for membership in flagella clusters (c11 to c14) versus other clusters (c1 to 10, c15 to c18, and unclustered) and for overlapping membership between the three studies. Resampling without replacement from 10,000 replicates (as described in the preceding section) was used to derive expected null distributions in each overlap category and compared with actual data derived by separating c11 to c14 genes in each study from the remaining genes (Figure 6B; Supplemental Figure 15).

Coexpression of Paralogs Encoding Enzymes of Central Carbon Metabolism

Pathway signature genes, defined as genes encoding enzymes whose activity is unique to a specific metabolic pathway, were identified as follows: glyoxylate cycle, MAS1 and ICL1 (encoding malate synthase and isocitrate lyase); TCA cycle, SDH3 and SDH4 (encoding succinate dehydrogenase subunits 3 and 4); CBB cycle, RBCS1 and SBP1 (encoding Rubisco small subunit 1 and sedoheptulose-1,7-bisphosphatase); fermentation, PFL1 (encoding pyruvate-formate lyase); pentose phosphate pathway, GLD1 and GLD2 (encoding glucose-6-phosphate dehydrogenase); and the committed steps of glycolysis, PFK1 and PFK2 (encoding phosphofructokinase) and gluconeogenesis (encoding phosphoenolpyruvate carboxykinase and fructose-1,6-bisphosphatase). The expression estimates of these signature genes were used to determine expression profiles indicative of each pathway by averaging scaled expression values for the signature genes in a pathway. Pearson’s correlation coefficients were then calculated for each paralog that potentially participate in each of these pathways and each pathway expression profile to evaluate coexpression between pathway-containing genes. Paralogs were assigned to pathways on the basis of highest correlation to the signature genes, using a correlation threshold } 0.8.

Curation of a Light Stress Response Cluster

A custom light stress response cluster of 280 genes was generated whose members came almost entirely from a subset of genes in c1 and c2 that showed a sharp transient peak of expression at ZT1. Criteria for inclusion were an expression maximum at ZT1 that was > 5 RPKM, expression levels at ZT3, ZT4, and ZT5 no greater than 80, 60, and 50% of the maximum, and expression at ZT6-ZT24 no higher than 40% of the maximum.

Gene Expression Graph Smoothing

Gene expression graphs in Figure 3 were smoothed by a custom rolling average function created within the programming shell R (http://www.R-project.org/). Absolute expression estimates were first averaged and then normalized to maximum expression. The estimates were then converted to the weighted average of samples taken within 1 h of each time point with the central sample given full weight, those 1 h from the central sample given quarter weight, and those 30 min from the central sample given half weight. Samples ZT1 and ZT24 were each used for their respected transformations.

Gene expression graphs in Supplemental Figures 20 and 21 were smoothed using the FFT method of signal processing with two points in Origin (OriginLab).

Accession Numbers

Primary data are available at the NCBI Gene Expression Omnibus repository under accession number GSE71469. Chlamydomonas locus ID numbers for genes described in this study are listed in the appropriate supplemental data sets and can be accessed from Phytozome (http://phytozome.jgi.doe.gov).

Supplemental Data

Supplemental Figure 1. Comparison of diurnally cycling genes identified with JTK cycle and DESeq2.
Supplemental Figure 2. Expression of rhythmic genes that were identified by DE analysis but not by JTK cycle.
Supplemental Figure 3. Expression of rhythmic genes that were identified by JTK cycle but not by DE analysis.
Supplemental Figure 4. Expression profiles of DNA replication genes I.
Supplemental Figure 5. Expression profiles of DNA replication genes II.
Supplemental Figure 6. Expression profiles of SMC genes.
Supplemental Figure 7. Expression profiles of cell cycle regulatory genes.
Supplemental Figure 8. Expression profiles of RDP genes.
Supplemental Figure 9. Expression profiles of anaphase promoting complex/cyclosome (APC/C) genes.
Supplemental Figure 10. Expression profiles of chloroplast division genes.
Supplemental Figure 11. Cell cycle gene resampling statistics.
Supplemental Figure 12. Expression patterns of DIV and GEX genes.
Supplemental Figure 13. Resorption and reformation of flagella during the cell cycle.
Supplemental Figure 14. Flagella and basal body gene cluster membership statistics.
Supplemental Figure 15. Resampling statistics for comparisons of flagellar gene group membership overlap.
Supplemental Figure 16. Ribosomal protein gene transcript contribution to the transcriptome.
Supplemental Figure 17. Expression profiles of nuclear genes encoding subunits of photosynthetic complexes.
Supplemental Figure 18. Expression profiles of genes encoding subunits of complexes I-III of the mitochondrial electron transport chain.
Supplemental Figure 19. Expression profiles of genes encoding complexes IV-V of the mitochondrial electron transport chain.
Supplemental Figure 20. Expression profiles of nuclear genes encoding subunits of chloroplast and mitochondrial ATP synthase complexes.
Supplemental Figure 21. Expression profiles of genes encoding enzymes of tetrapyrrole metabolism.
Supplemental Figure 22. Expression profiles of genes encoding LHC-like proteins.
Supplemental Figure 23. Expression profiles of transcription factor genes.

Supplemental Table 1. Illumina read mapping statistics.

Supplemental Data Set 1. Gene expression estimates, cluster membership, and additional data.

Supplemental Data Set 2. Genes with constant expression.

Supplemental Data Set 3. MapMan term cluster distributions and statistical enrichment tests.


Supplemental Data Set 5. DIV and GEX genes.


Supplemental Data Set 7. Flagella gene group membership.

Supplemental Data Set 8. Ribosomal gene expression estimates.


Supplemental Data Set 10. Respiratory complex gene expression.

Supplemental Data Set 11. Tetrapyrrole metabolism-related gene expression.

Supplemental Data Set 12. Light stress cluster gene expression.

Supplemental Data Set 13. LHC-like gene expression.


Supplemental Data Set 15. Central carbon metabolism enzyme gene expression and pathway assignments.

Supplemental Data Set 16. Transcription factor gene expression.

Supplemental Data Set 17. Nonexpressed genes.

ACKNOWLEDGMENTS

This work was supported by National Institutes of Health Grant R01 GM069521 to J.G.U. and by the University of California-Los Angeles Department of Energy Institute of Genomics and Proteomics (DE-FC03-02ER63421) and National Institutes of Health (NIH) R24 GM092473 to S.S.M. J.M.Z. was supported by the National Science Foundation Integrative Graduate Education and Research Traineeship Fellowship (0504645). I.K.B. was supported by a training grant from the NIH (T32ES015457) and the Graduate Education and Research Traineeship Fellowship (0504645).


AUTHOR CONTRIBUTIONS


REFERENCES


Received June 8, 2015; revised July 27, 2015; accepted September 14, 2015; published October 2, 2015.


