The diurnal transcriptional landscape of the microalga Tetradesmus obliquus

Carreres, B. M., Mitsue León-Saiki, G., Schaap, P. J., Remmers, I. M., van der Veen, D., Martins dos Santos, V. A. P., ... Suarez-Diez, M.

The diurnal transcriptional landscape of the microalga *Tetradesmus obliquus*

Benoit M. Carreres [1], G. Mitsue León-Saiki [2], Peter J. Schaap [1], Ilse M. Remmers [2], Douwe van der Veen [2], Vitor A.P. Martins dos Santos [1], René H. Wijffels [2,3], Dirk E. Martens [2+], Maria Suarez-Diez [1+]

[1]Laboratory of Systems and Synthetic biology, Wageningen University & Research, the Netherlands

[2]Bioprocess Engineering, Wageningen University & Research, the Netherlands

[3]Nord University, Faculty of Biosciences and Aquaculture, N-8049, Bodø, Norway

+ corresponding authors. These authors jointly supervised the work.

Abstract

Tetradesmus obliquus is a promising oleaginous microalga. We functionally annotated its genome and characterized the transcriptional landscape of *T. obliquus* adapted to 16:8h light dark (LD) cycles in turbidostat culture conditions at very high temporal resolution (1h intervals). Revealing a cycle of cellular events, six distinct expression profiles were obtained, each with transcriptional phenotypes correlating with measurements of biochemical composition.

The impact of starch deficiency was studied using the starchless mutant *slm1*. Significant changes in the transcriptional landscape were observed. Starch deficiency resulted in incapacity to supply energy during the dark period, resulting in a shift of energy demanding processes to an earlier or later time point. Our study provides new perspectives on the role of starch and the adaptation to LD cycles of oleaginous microalgae.

Keyworkds: Microalgae; *Scenedesmus obliquus;* starchless mutant; diurnal transcription changes; photosynthetic efficiency; day/night cycles

1 Introduction

Microalgae are a promising source of compounds of interest (lipids, proteins, and pigments) for the production of food, feed, chemicals, and fuels [1–3]. Large scale microalgal production will be primarily done outdoors under natural diurnal light/dark (LD) cycles [4,5]. Diurnal cycles are ubiquitous and photosynthetic organisms synchronize their metabolic activities to anticipate light changes in the environment and schedule specific tasks during the night or day [6–9]. Synchronization in photosynthetic organisms involves the regulation of photosynthesis to maximize carbon fixation and use of light during the day and to schedule light sensitive processes (such as DNA synthesis and cell division) at night [6, 10–12].

Tetradesmus obliquus is a microalga recognized as an industrially relevant strain for food and fuel production [13, 14]. *T. obliquus* can reach a maximum triacylglycerides (TAG) content of 0.45 $g \cdot g_{DW}^{-1}$ and a maximum TAG yield on light of 0.14$g \cdot (mol\ photons)^{-1}$ under batch nitrogen starvation and continuous light conditions [15]. For further improvement of TAG yield on light, de Jaeger et al. developed the starchless mutant *slm1* [16], which cannot synthesize starch due to a missense mutation in the small subunit of ADP-glucose pyrophosphorylase, the committed step of starch biosynthesis [17]. Under culture conditions of continuous light and batch nitrogen starvation, *slm1* showed a higher maximum TAG yield on light (0.217 $g \cdot (mol\ photons)^{-1}$) and maximum TAG content (0.57 $g \cdot g_{DW}^{-1}$) compared to the wild-type (WT) without a decrease in photosynthetic efficiency [15]. Comparable results were obtained under light-dark cycles and batch nitrogen starvation [14].

LD cycles give an advantage over continuous light, leading to higher energy conversion efficiency and higher yield of biomass on light. The physiological behavior of *T. obliquus* WT and of the starchless mutant (*slm1)* was also studied under 16:8h LD cycles in turbidostat controlled systems [8,10]. Under this light regime and nitrogen replete conditions, *T. obliquus* showed synchronization and diurnal patterns of its metabolism, which amongst others, suggested that starch was used as a temporary energy storage. This is,

starch was accumulated during the light period and was used during the dark period. Cell division was also synchronized and occurred mainly during the night. The starchless mutant *slm1* showed a lower energy conversion efficiency (11-24% lower) and biomass yield on light (13-39% lower) compared to the WT under different photoperiods [10]. Furthermore, for the *slm1* mutant cell division still occurred mainly during the night, but at a slower rate and no diurnal oscillations on any of the other measured compounds were found [10]. Unlike for the WT, diurnal LD cycles did not provide an advantage for the *slm1* mutant compared to continuous light. On light, biomass yield as well as the energy conversion efficiency were similar [8]. Starch may therefore play an important role in harvesting additional light energy during LD cycles.

While the biochemical analysis allowed us to draw some conclusions, we still lack information on how cellular processes are regulated during the diurnal cycle and how the synthesis and use of starch is connected to these processes. Furthermore, while the inability to make starch had an impact on energy conversion efficiency, it is not known how this will affect the timing and regulation of the different cellular processes. LD cycles and the subsequent synchronization of the microalgal population makes it of paramount importance to unravel the timing of cellular and subcellular events. Understanding the timing at which metabolic changes take place is essential to understand the phenotype exhibited by the WT and *slm1* strains and to optimally design experiments characterizing mutant strains.

Time resolved transcriptome analysis can give useful insights into the timing and regulation of the different physiological stages and on the succession of processes in the cell. Thus it allows the association of the cellular processes with biochemical properties. Algal diurnal transcriptional regulation is not well known. To our knowledge, only a few reports on diurnal oscillations under light/dark cycles in microalgae have been published [18–23] The first one was on the eukaryotic red alga *Cyanidioschyzon merolae* [24] grown under 12:12h LD cycles, where they studied the transcriptional changes in intervals of 2 h. Furthermore, the diurnal transcriptional changes were studied in the diatom *Phaeodactylum tricornutum* [20] under 16:8h LD cycles. The authors studied changes in biochemical composition (carbohydrates and lipids) in 8 different unequally distributed time points during a period of 26.5 hours (5 points during the light period and 3 during the dark period). Finally, transcriptional analysis on *Nannochloropsis oceanica* [21] and *Chlamydomonas reinhardtii*

[22] under a 12:12h LD cycle was published. For *N. oceanica* the authors studied growth, changes in biomass composition (lipids and glucose) and in gene expression in intervals of 3h, while for *C. reinhardtii* a more detailed study was done with intervals of 1h during most of the cycle and every 30 min for some time points, ending up with a total of 28 points distributed over the 24h cycle. Our study, however, is the first one to look into the transcriptional changes in a diurnal cycle of an oleaginous green algae with a high sampling frequency (intervals of 1 h). Additionally, this is the first study reporting on the transcriptome changes over a diurnal cycle for a starchless mutant, which could give insights into the role of starch in microalgae.

To fully characterize the changes induced by LD cycles and the role of starch metabolism, we analyzed and compared the transcriptional landscape of *T. obliquus* WT and *slm1* under diurnal 16:8h LD cycles. For this, we used the same samples collected from our previously published article where also the measurements of the biochemical composition are published [8]. Each strain was cultivated in two separate photobioreactors that were operated in a continuous turbidostat controlled mode under a 16:8 LD cycle resulting in an oscillating steady state that was synchronized to the light regime. Samples from all four photobioreactors were taken in intervals of 1h for WT, and intervals of 3h for *slm1*. Finally, changes in the transcriptional landscape were related with their previously published [8] measurements of biochemical composition.

2 Materials and methods

2.1 Experimental setup and sampling

Wild-type (WT) *Tetradesmus obliquus* UTEX 393 (formerly known as *Acutodesmus obliquus and Scenedesmus obliquus* [25]) was obtained from the Culture Collection of Algae, University of Texas. The starchless mutant of *T. obliquus* (*slm1*) was generated as described by de Jaeger et al. [16]. *T. obliquus* was continuously cultivated in a sterile flat panel airlift-loop reactor with a 1.7 L working volume (Labfors 5 Lux, Infors HT, Switzerland). Culture conditions and reactor set-up (27.5 °C, pH 7.0 and gas flow rate of 1 $L \cdot min^{-1}$ air enriched with 2% CO_2) were controlled as previously described in León-Saiki et al., where also the measurements of the biochemical composition are published [8]. Light was provided in a 16:8h light/dark

(LD) block at an incident photon flux density of 500µmol·m^{-2}·s^{-1} (warm white spectrum 450-620 nm). Cultivations were turbidostat controlled, where fresh medium was fed to the cultures when the light intensity at the rear of the reactor dropped below the setpoint (10µmol·m^{-2}·s^{-1}, OD_{750}). Feeding of medium was stopped during the dark period. After steady state was reached, 8mL samples were taken for RNA extraction (approximately 10 mg_{DW}). Cells were immediately collected by centrifugation (4255 xg, 0°C for 5 min), supernatant was discarded and pellets were frozen in liquid nitrogen and stored at -80°C until further extraction. Samples for RNA extraction were taken in intervals of 1h for WT or every 3h for *slm1*. Due to restrictions on working hours of the laboratory, the samples were collected in two successive time settings to allow sampling the dark period during the day. After collecting samples of the first half of the cycle, light settings were shifted and the culture was then allowed to reach oscillating steady-state before collecting samples for the second half of the cycle. The first and the last samples of each time settings are overlapping samples for control. Therefore, four RNA samples are present at these time points. Overall 72 samples were taken for RNA extraction.

2.2 RNA isolation and quality control

RNA extraction was performed using the Maxwell® 16 LEV simplyRNA Tissue kit (Promega). Frozen algae pellet (≈200 µL) were submerged in 400 µL of homogenizing buffer supplemented with 8 µL 1-thioglycerol in a 2 mL Lysing matrix C tube (MP), prefilled with a mix of glass beads. Samples were disrupted using a FastPrep-24 instrument (MP). After disruption, all liquid was transferred to a LEV RNA Cartridge. 300 µL lysis buffer (from the kit) were added and the rest of the extraction was performed using a Maxwell MDx AS3000 machine (Promega) following manufacturer's instructions. RNA integrity and quantity were assessed with an Experion system (Bio-Rad), and only high quality samples (RIN value ≥ 7) were selected. Total RNA was sent for whole transcriptome sequencing to Novogene Bioinformatics Technology Co. Ltd (HongKong, China).

The corresponding data have been submitted to EBI ArrayExpress and can be found under the accession number E-MTAB-7009.

2.3 Genome structural annotation

Using the available genome sequence of *T. obliquus* [26],we performed an RNA-Seq-based genome annotation using BRAKER1 [27]. RNA reads from 38 samples of both WT and *slm1*, supplemented with an additional sample from each strain under nitrogen limited condition were given as additional BRAKER1 input. The annotation information was processed using our semantic framework pipeline [28] and the information was stored according to our integrated ontology [29] which respects the FAIR data principles [30]. The annotation framework and the ontology were extended with the necessary tools [27,31,32] and ontology terms for the purpose of this analysis.

The genome feature annotation has been used to updated the original genome ENA project with accession number PRJEB15865.

2.4 Genome functional annotation and pathway mapping

Proteins were annotated to GO terms using InterProScan5 and Argot2, with default parameters [33,34]. Proteins were annotated to enzyme commission (EC) number using EnzDP [35]. Results with a complete EC number and a likelihood-score of at least 0.2 were used for subsequent analysis. To choose a threshold for the likelihood-score that results in a good trade-off between true positive and false positives, we visually compared the completeness of KEGG metabolic maps while avoiding dispersion of each reaction into different expression clusters. The functional annotation can be found in supplementary file S1 that includes EC numbers and GO terms.

2.5 RNA-seq normalization and expression calculation

Using all the transcripts found from the genome annotation, we aligned the reads from each sample and calculated the FPKM values using Cufflinks [32]. As a pre-filter to remove unexpressed genes and false positives from gene annotation, we selected the transcripts with a coverage of at least 1 and a FPKM of at least 0.1 in at least 10 of the 113 analyzed samples. These samples include the 72 of this study, and the rest

were taken under LD cycles and nitrogen limited conditions (not studied in this manuscript). Additionally, we selected the transcripts having an expression of at least 10 samples with a value higher or equal to the 0.15 quantile. Following the advice of maSigPro [36], the FPKM of all samples were normalized using the scaling normalization method TMM [37] using the R functions from edgeR package "calcNormFactors" [38].

2.6 Gene clustering

To identify genes with significant expression profile changes over time, maSigPro was used [36,39], with the modified parameters: regression model was set to a maximum of 23 degrees, the parameter "counts" was set to true, "nvar.correction" was also set to true, the "step method" set as backward, and R-squared was set to 0.7. The R-squared was chosen based on the relevance and the balance in the number of enriched pathways identified in each cluster (see below). To better estimate the number of clusters to group the gene expression, we evaluated combinations of significance level (Q: 0.005 to 0.05) and R-squared (0.85 to 0.5). We decided to keep the standard-strict values (Q=0.05, R-squared=0.7), which would also give a good balance in pathway enrichment and number of genes in each group.

Hierarchical clustering with agglomerative linkage was performed using the R stats package (hclust function). Number of clusters from 3 to 25 were evaluated using the R function "cluster.stats" from package "fpc" [40]. For each set of clusters, the following indexes were computed: average silhouette widths, normalized gamma, two Dunn indexes, average within and average between ratio, and Calinski and Harabasz index. Additionally, each set of clusters were compared one on one with the Rand index.

2.7 Enrichment analyses

Enrichment analyses were performed using the hypergeometric function to model the probability density using the "phyper" function from the R package stats [41]. Two types of analysis were performed: pathway and GO term enrichment. Pathway enrichment required associating annotated Enzyme Commission (EC) numbers to metabolic maps. We used the online available resource from KEGG pathway maps [42, 43]. The

KEGG pathways fitting the following requirements were kept for further analysis: 60% coverage if 3 to 6 EC numbers annotated, and 50% coverage if 6 to 10 EC numbers annotated, and 25% of coverage if more than 10 EC numbers were annotated. For the hypergeometric test we considered the universe size, *N*, to be the total number of EC numbers in all pathways in the genome, *m* is the number of successes in this universe and is defined as the number of EC numbers in the corresponding pathway in the genome, *k* and *x* are the sample size and the number of successes in the sample (or considered gene subset) respectively. Enrichments with a p-value lower than 0.05 were considered significant. Similarly, for the GO enrichment, *N* is the total number of genes annotated to any GO terms in the genome, *m* is defined by the number of genes annotated to the considered GO term in the genome, and *k* and *x* refer to the considered subset of genes. Multiple test correction for the GO enrichment was performed using the Benjamini–Hochberg procedure. Enrichments with FDR<0.05 were considered significant. To handle the GO information such as ontology, ancestor, and offspring, we used “GO.db” database from Bioconductor [44]. Additionally, to reduce the number of GO terms and conserve the most specific, only terms not having any offspring in the selection were retained. We additionally provide the pathway coverage, and the *x* and *m* values to better understand the reason of each significant enrichment.

3 Results

3.1 Genome annotation

The genome of *T. obliquus* UTEX 393 is about 100 million base pairs in size [26]. BRAKER1 annotation revealed 19795 genes, which transcribed 21493 coding sequences and translates into 19723 protein sequences.

To analyze the changes in transcription over a diurnal cycle of *T. obliquus* WT and *slm1* strains, we sequenced RNA from biological duplicate runs. WT was sampled every hour in duplicate resulting in 52 samples. *slm1* was sampled in intervals of three hours resulting in 20 samples. For both strains, time points 0h and 13h were overlapping samples and therefore sequenced in quadruplicate (see material and methods

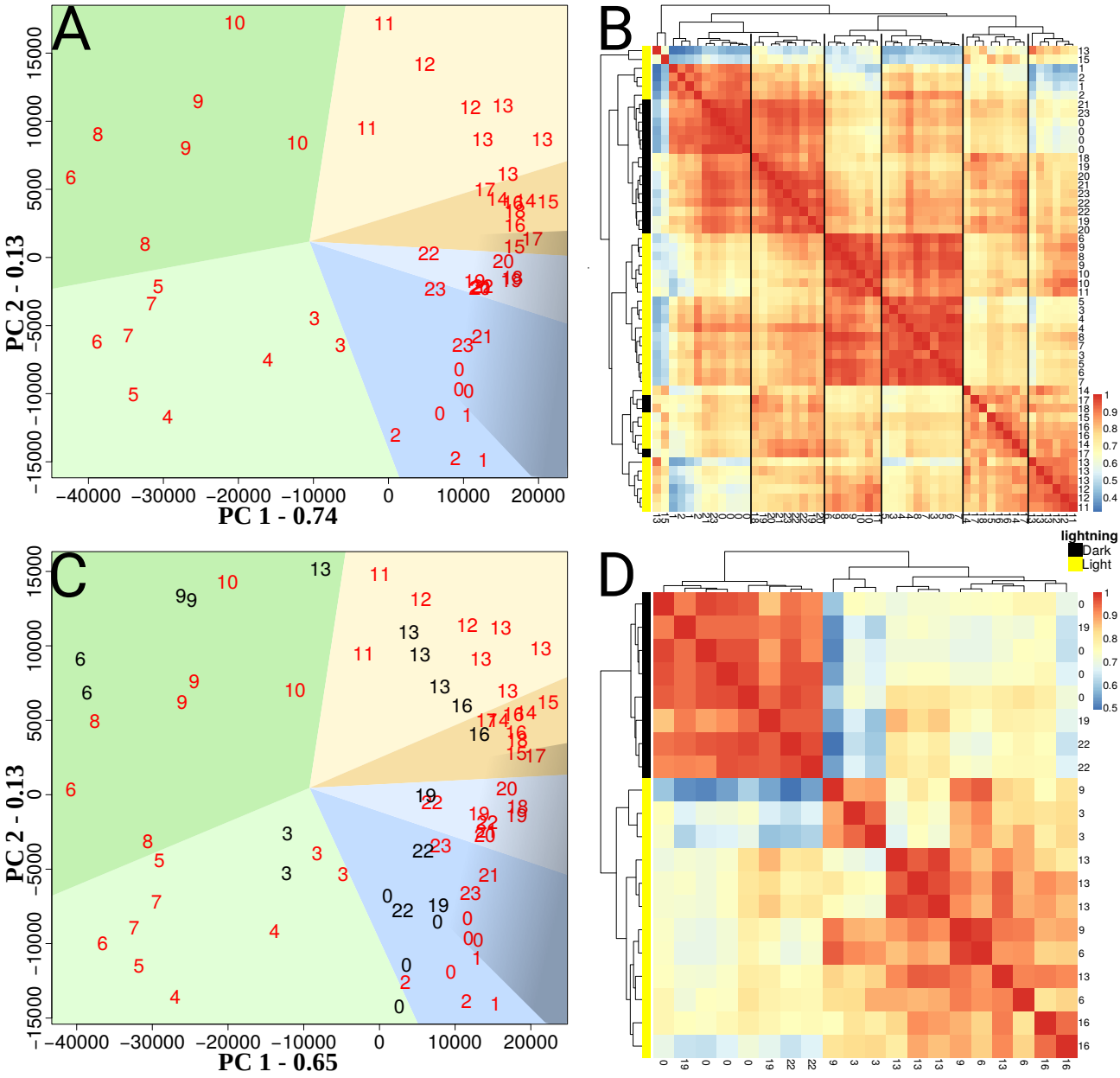
for details). Not all the transcripts were found to be expressed and filtering was performed to identify genes with very low expression that were not further considered. From the 19723 proteins, 16810 proteins remained after filtering for subsequent analysis. GO term annotation yielded 13687 proteins annotated to 2394 unique GO terms. EC annotation yielded 3559 proteins annotated to 1315 EC numbers. The functional annotation can be found in supplementary file S1.

3.2 Light-dark 16:8h cycle induces systemic transcriptional changes following a circular pattern

4686 genes were found to have a significant change of expression during the diurnal cycle. A principal component (PC) analysis (PCA) for these selected genes is shown in Figure 1A. The PC plot provides a global overview of changes in gene expression over time. The time point samples form an ordered and circular pattern over the two first PCs. These two first PCs together explain 87% of the variation in expression. To demonstrate that this is not a consequence of gene selection, we rendered another PCA considering the whole set of expressed genes (supplementary file S2). Within this set, the overall pattern is noisier, but the two first components still explain 78% of the variation in expression. The main changes in gene expression occur along PC1 from 1h to about 5h and back again from 8h to about 13h. The second important change occurs along PC2 from 6h to 9h and back again from about 22h to 0 h. During the period from 14h to 22h there is very little change in gene expression. Overall, genes associated to light harvesting complex (LHC) were found to be the main contributors of PC1. The two highest contributing genes to PC2 are glyceraldehyde-3-phosphate dehydrogenase (GA3PDH) and glucose-6-phosphate isomerase (GPI). LHC genes also contributed strongly to PC2.

The heatmap representation in Figure 1B offers a complementary view of these multi-dimensional data. It allows to evaluate differences and similarities in expression between the time points. Two samples corresponding to time points 13h and 15h are separated in the dendrogram from the neighboring samples and from the biological duplicate samples. However, in the heatmap it can be seen that these samples still show

relatively high correlation with their corresponding biological duplicates and samples from the surrounding hours. Three main groups of time points can be identified in the dendrograms in Figure 1B, which we will refer to as time phases I, II, and III. Additionally, each phase can be subdivided in two sub-phases, a and b, which represent subtler expression changes. The occurrence of the phases over the 16:8h LD cycle is shown in the Figure 1E for both strains. The phases were named according to their order of appearance from 0h to 23h. These time phases are also apparent and can be associated to changes along the two first PC. Similarly, time phases were also associated to the *slm1* strain based on the PCA plots. For *slm1*, we noticed a change in phase timing that represents a significant change of expression occurring one to two hours before each dark-light shift.



E	Time (h)	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
	Light		ON																OFF						
Phase	WT	Ia			IIa					IIb			IIIa			IIIb				Ib				Ia
	MT	Ia			IIa			IIb			IIb			IIIa			IIIb				Ib			Ia	

Figure 1: *Principal Component Analysis (PCA) (A,C) of gene expression data for* Tetradesmus obliquus *wild-type (WT) (A) and both WT and starchless mutant (*slm1*) (C) during 16:8h light-dark cycle. Numbers represent the time points of the samples, with red for the WT samples and black for the* slm1 *samples. Background colors in both PCA plots refer to the colors given to the six time sub-phases described in (E), and the dark shade refers to the dark period. (B, D) Heatmaps showing similarity of gene expression data among samples for WT (B) and* slm1 *(D) strains, samples are identified by their time points. Dark vertical lines across the heatmap in (B) indicate different time time phases. (E) Overview of the time phases.*

3.3 Gene expression clusters and their time profiles

3.3.1 Cluster profiles

We clustered the 4686 genes that showed significant changes over time in different numbers of clusters ranging from 3 to 25. Next we evaluated the cluster separation using seven well established indexes to assess similarities within clusters and differences between clusters. These results are depicted in the supplementary file S3. Some of the selected metrics (such as the first Dunn index) gave no clear indication on the optimal number of clusters. However other metrics such as average silhouette width, second Dunn index, normalized gamma or Calinski-Harabasz index, show a local maximum (or minimum in the case of the average within and average between ratio) at 4-6 clusters. Inspection of cluster similarities based on the Rand index (supplementary file S3) led to select 6 clusters (over 4 or 5) as it would result in a new and distinct cluster. Based on these results, we considered separating the genes in six clusters to be optimal.

The resulting six cluster profiles were summarized using for each time point the median value of gene expression of each gene present in the cluster and are depicted in Figure 2. Cluster 1 to 6 contain respectively 829, 364, 1020, 929, 952 and 592 genes. All the cluster profiles are different, but there are certain similarities between clusters 1 and 6, and between 3, 4, and 5. Interestingly, the two peaks of expression observed in cluster 3, 4, and 5 are all 5-6 hours apart.

The temporal profile of each cluster peaks at different moments of the diurnal cycle. Cluster 1 peaks the earliest, 1h after light goes on. Cluster 2 peaks at 3h, cluster 3 peaks around 11h and shows a smaller secondary peak at 17h. Cluster 4, with a similar profile, shows the first peak at 13h and the second at 18h. Cluster 5 has a rather gradual increase of expression and peaks around 14h followed by a smaller peak at 20h. Finally, cluster 6 has very gradual expression changes throughout the day, and shows a peak at 23h.

3.3.2 Cluster functional analyses

To understand the functions of the genes in the clusters, we performed two sets of complementary enrichment analyses. Metabolic pathway enrichment analysis provides direct information on gene metabolic functions. The functional annotation identified genes associated to EC numbers. Mapping these EC numbers to KEGG pathway maps revealed 67 pathways for which enough genes can be associated to the corresponding reactions to warrant further analysis, as detailed in Materials and methods. GO enrichment analyses were performed on the three ontologies: biological process (BP), molecular function (MF) and cellular component (CC). These analyses provide a wider overview not restricted to metabolism. The enrichment results are summarized in Table 1. Each cluster median time profiles and the full set of significant (p-value < 0.05) enrichments are presented in the supplementary file S4.

Cluster 1 shows enrichment in pathways related to amino acids (AA), and riboflavin metabolism. GO terms enrichment is in agreement and contains terms such as tRNA modification, aminoacyl-tRNA biosynthesis, ribosome, alpha-amino acid biosynthetic process and translation. All these pathways and terms indicate protein synthesis and the associated strong demand for AA. Riboflavin metabolism, leading to FAD synthesis, and ubiquinone and other terpenoid-quinone biosynthesis appears enriched in cluster 1. Inspection of this last pathway (supplementary file S5) indicates that ubiquinone is the most probable product in this time frame. Ubiquinone is an important step in the oxidative phosphorilation by oxidating $FADH_2$ back into FAD and it indirectly plays a role in ATP synthesis.

Cluster 2 mainly shows enrichment in the processes related to pigments synthesis, which include carotenoids, chlorophylls, and their precursor (Phytyl-diphosphate) from the terpenoids backbone synthesis. Those are needed for building the photosystems and starting the carbon fixation. Three of the four reactions involved in the starch synthesis were found in cluster 2, this can be observed from the pathway map (supplementary file S5). Cluster 2, also shows enrichment in genes related to transcription, amino acid (AA) and protein synthesis, but to a lower extent than cluster 1. The carbon fixation pathway is found enriched in cluster 2, which goes in line with enrichment in the pigments and in starch synthesis. The GO terms in

biological process are in agreement with the pathway enrichments. The GO terms in molecular function provide extra information on the nature of chemical reactions performed by these genes. We also found that this cluster contains genes associated to the LHC that were found to strongly contribute to PC1 and PC2 and among the proteins associated to carbon fixation in this cluster GAPDH was found to be a main contributor in PC2. Interestingly, while all the other amino-acid synthesis pathways were found enriched in cluster 1, the pathway of “Phenylalanine, tyrosine and tryptophan biosynthesis” is found enriched in cluster 2. This result indicates that these aromatic amino-acids are probably synthesized later than the other, more simple, amino-acids.

In cluster 3, the GO enrichment provides valuable information that cannot be covered by the pathway enrichment. The GO term enrichments display clear terms such as “DNA replication”, “organelle fission”, “chromosome organization”, “microtubule”, “cytoskeletonDNA-directed”, “DNA polymerase activity”. This cluster clearly groups all processes related to the full cell cycle. The pathway enrichment reveal processes related to AA pathways: “Lysine degradation” and “Phenylalanine metabolism”. Additionally, the enrichment in “N-Glycan biosynthesis” and “Various types of N-glycan biosynthesis” together with the enriched GO terms “protein complex assembly” and “vesicle-mediated transport” indicates important protein maturation processes leading to complex proteins and some transport of proteins to membranes.

Cluster 4 is mostly enriched in pathways associated to different kind of carbohydrates metabolism. The first pathways enriched are “Galactose metabolism“, “Starch and sucrose metabolism”, “Other glycan degradation”, and “Amino sugar and nucleotide sugar metabolism”. With the GO terms displaying “polysaccharide binding”, “hydrolase activity”, and “hydrolyzing O-glycosyl”, we observed the same trend towards starch degradation. The GO term for “microtubule binding” is the only one of this kind in this cluster, but it comes right after cluster 3 where a lot of cytoskeletal and microtubule terms were found enriched. Finally, there is enrichment in Cytochrome P450 which hints on repair mechanisms potentially related to photo-damage.

Cluster 5 shows enrichment in a very diverse set of metabolic pathways, covering glycolysis, pyruvate metabolism, glutathione metabolism, co-enzyme-A, starch and other carbohydrates polymers. More

importantly, the cluster is also enriched in genes related to nitrogen metabolism and glutamate associated genes. ATP synthase is found in this cluster too, but the gene is highly expressed throughout the whole light period with highest expression before dark. Among the genes associated to the chlorophyllase (EC 3.1.1.14), the best candidate (g19660.t1) is found in cluster 5. This also agrees with the observed enrichment of Cytochrome P450, which was recently found to take major part in the breakdown of chlorophyll in *Arabidopsis thaliana* [45–47].

Cluster 6 shows enrichment in fatty acids degradation and folate biosynthesis and contains the reactions from GTP until the precursor of dihydrofolate (DHF).

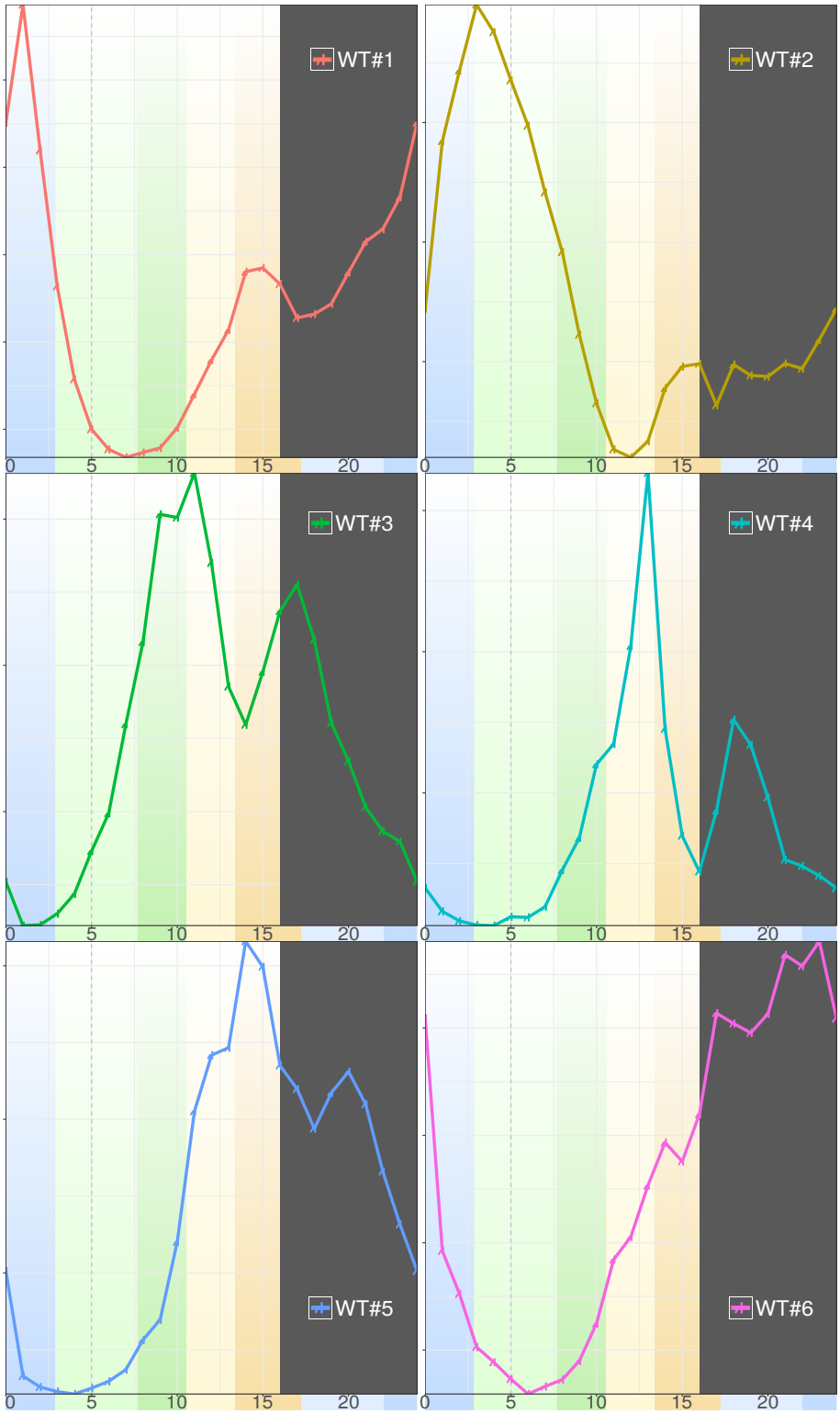


Figure 2: *Expression profile of the 6 gene clusters for* Tetradesmus obliquus *under 16:8h light-dark cycles. The plots show the median profile of gene expression in the indicated clusters. The background colors correspond to the colors given to the time phases in Figure 1. The vertical gray dashed line represents the time point of maximum dilution rate taken from [8]. Dark area represents the dark period.*

Table 1: *Summary of the results of the enrichment analyses. The first column contains the cluster identifier. The second contains the enriched pathways (p-value < 0.0.5). The last three columns are the results of the GO enrichment analyses (FDR<0.05) for each of the GO ontologies: biological process (BP), cellular component (CC) and molecular function (MF). Full set of GO enrichment results are available in the supplementary file S4.*

Cluster	Pathways	BP	CC	MF
1	Ubiquinone and other terpenoid-quinone biosynthesis Purine metabolism Glycine, serine and threonine metabolism Cysteine and methionine metabolism Valine, leucine and isoleucine biosynthesis Lysine biosynthesis Arginine and proline metabolism Riboflavin metabolism Aminoacyl-tRNA biosynthesis	RNA methylation tRNA modification translation nucleobase-containing small molecule metabolic process alpha-amino acid biosynthetic process	ribosome	RNA binding structural constituent of ribosome GTPase activity GTP binding S-adenosylmethionine-dependent methyltransferase activity ligase activity, forming carbon-nitrogen bonds
2	Glycine, serine and threonine metabolism Phenylalanine, tyrosine and tryptophan biosynthesis Carbon fixation in photosynthetic organisms Porphyrin and chlorophyll metabolism Terpenoid backbone biosynthesis Carotenoid biosynthesis Aminoacyl-tRNA biosynthesis	tRNA aminoacylation for protein translation steroid biosynthetic process coenzyme metabolic process porphyrin-containing compound biosynthetic process photosynthesis pigment biosynthetic process oxidation-reduction process	photosystem	nucleotide binding 3-beta-hydroxy-delta5-steroid dehydrogenase activity aminoacyl-tRNA ligase activity iron ion binding calcium ion binding oxidoreductase activity, acting on a sulfur group of donors oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen transferase activity, transferring alkyl or aryl (other than methyl) groups carbon-carbon lyase activity hydro-lyase activity heme binding precorrin-2 dehydrogenase activity coenzyme binding 2 iron, 2 sulfur cluster binding
3	Lysine degradation Phenylalanine metabolism Glutathione metabolism N-Glycan biosynthesis Various types of N-glycan biosynthesis Propanoate metabolism	DNA replication DNA repair DNA recombination protein complex assembly vesicle-mediated transport cell cycle process regulation of cellular metabolic process proteasome-mediated ubiquitin-dependent protein catabolic process organelle fission chromosome organization regulation of primary metabolic process carbohydrate derivative biosynthetic process single-organism organelle organization	microtubule cytoskeleton protein complex cytoskeletal part	DNA binding DNA-directed DNA polymerase activity ATP binding microtubule binding transcription factor binding four-way junction helicase activity ATPase activity, coupled
4	Galactose metabolism Starch and sucrose metabolism Other glycan degradation Amino sugar and nucleotide sugar metabolism Metabolism of xenobiotics by cytochrome P450 Drug metabolism - cytochrome P450	carbohydrate metabolic process		hydrolase activity, hydrolyzing O-glycosyl compounds microtubule binding polysaccharide binding
5	Glycolysis / Gluconeogenesis Glutathione metabolism Starch and sucrose metabolism Sphingolipid metabolism Pyruvate metabolism Glyoxylate and dicarboxylate metabolism Thiamine metabolism Pantothenate and CoA biosynthesis Nitrogen metabolism	carbohydrate metabolic process oxidation-reduction process		oxidoreductase activity, acting on CH-OH group of donors

	Metabolism of xenobiotics by cytochrome P450 Drug metabolism - cytochrome P450			
6	Fatty acid degradation Pyrimidine metabolism Butanoate metabolism Folate biosynthesis			

3.4 Transcriptional landscape of *Tetradesmus obliquus slm1.*

We compared changes in gene expression between the starchless mutant (*slm1*) and WT to understand the differences resulting from the lack of starch accumulation. An overview of gene expression data in *slm1* in comparison to WT is shown in Figure 1C. The PCA shows very similar regulation over time in both strains. While many time points display similar expression patterns, there are also clear differences related to the separation in time phases. At the time points from 16h until 3h (phase IIIa to phase Ia), all *slm1* samples along PC1 are shifted left compared to WT, but they remain at the same level along PC2. This reflects a state of expression of WT that is never reached by slm1, rather than a diurnal dysregulation. On the contrary, the samples at 6h and 9h along PC2 are shifted up for *slm1* as compared to WT, but remain at the same level along PC1. This reflects an early state of expression, especially for *slm1* samples at 6h, which fits into the phase IIb. Another smaller time dysregulation is a small shift of time point 22h that is shifted down for slm1 as compared to WT. This reflects an early state of expression that fits into to the time phase Ia. The relative time phases of *slm1* are also shown in Figure 1.E. Overall, two main trends are observed: earlier changes in expression of processes shortly after light and a delay in change of expression of the processes before the dark period.

The PCA considering the whole set of expressed genes (Supplementary file S2) displays a very similar expression between WT and *slm1*. While there is a general overlap, the expression of *slm1* seems noisier, especially for the time points between 19h and 0h.

3.4.1 Comparison of the gene expression dynamics in WT and *slm1*

Using the regression based approach of maSigPro, we identified genes with differences in expression profile between *slm1* and WT. Our experimental design included sequencing samples of WT every hour and of *slm1* every three hours. However, maSigPro is designed to allow uneven distribution of time points. 784 genes showed no significant differences in expression profile between WT and *slm1* strains. 3902 genes showed significant differences in expression profile between WT and *slm1*. Additionally, 40 genes were found to have a time profile in the *slm1*, while no time profile was detected in WT for these genes. Apart form having no time profile it could also be that these genes are too noisy or are not expressed in WT.

3.4.2 Differences in time profiles

Clustering was done for the gene expression data obtained from the mutant *slm1*. Analysis of clustering performance indicators again resulted in 6 clusters. As previously stated, 784 genes did not show any significant difference in time dynamics between the two strains, therefore, these genes were used to guide the identification of the mutant clusters in relation to the wild type clusters. The cluster assignments of these 784 genes in both strains is shown in Figure 3A. The genes in WT clusters 3 and 4 are all found in the same *slm1* cluster, consequently named 3&4. Likewise, the WT genes in clusters 1 and 6 are all found in the same *slm1* cluster, consequently named 1&6. The WT genes in cluster 2 and 5 ended up in separate *slm1* clusters and consequently kept the same cluster name. Finally, the two remaining clusters, named A and B, do not contain genes with conserved expression. Those clusters have a time profile that does not fit any of the cluster profiles in WT.

Figure 3B shows the cluster assignments of the 3902 genes with significant differences in their time profiles when comparing WT and *slm1*. From these genes with altered expression (Figure 3B), there is a general trend of genes to remain clustered together. This means that most genes are transcriptionally controlled in large groups. The genes in WT cluster 3 are almost all found in the associated *slm1* gene cluster 3&4. Again,

almost all genes in WT cluster 2 are found in the associated *slm1* cluster 2. The genes in WT cluster 5 are distributed over five *slm1* clusters, with the majority in the associated *slm1* cluster 5. The genes in WT cluster 6 are distributed over three clusters, with the majority in the associated *slm1* cluster 1&6. Interestingly, a minority of genes in WT cluster 4 is found in the associated *slm1* cluster 3&4, but the majority is found in *slm1* cluster B. Similarly, a minority of WT genes in cluster 1 is distributed over three *slm1* gene clusters, with a small portion in the associated *slm1* cluster 1&6, a bigger portion in *slm1* cluster 1, and the majority in *slm1* cluster A. Thus the new cluster A mainly receives genes from the WT clusters 1, and to lesser extent from cluster 3, 5 and 6. On the other side, the new cluster B receives genes mainly from cluster 4 and to a lesser extent from cluster 5.

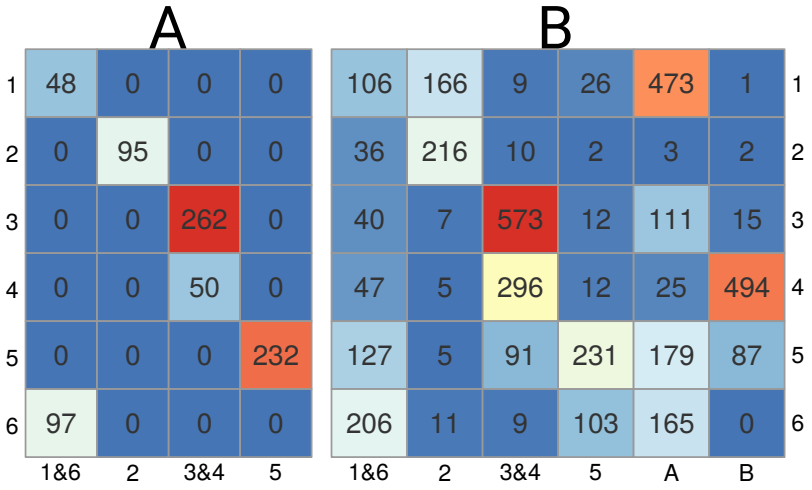


Figure 3: *Distribution of genes between time profile clusters of* Tetradesmus obliquus *wild-type (WT) (rows) and* slm1 *(columns). A: 784 genes with the same time expression profile in slm1 and WT. B: Genes with significantly different time profiles between* slm1 *and WT, as identified by maSigPro (3902 genes). Background color ranges from blue, white, to red, with red being the highest value and blue the lowest.*

slm1 cluster median profiles are depicted in Figure 4, together with the associated WT clusters. From cluster 1&6 to cluster B each *slm1* cluster grouped respectively 735, 507, 1304, 621, 958, and 600 genes. Overall, the time profiles of *slm1* clusters show strong similarities with the associated WT clusters. *slm1* clusters A and B are displayed together with WT clusters 1 and 4, respectively, due to the large number of genes found in common. The new cluster A is characterized by low expression at dark and the first half of the light period, with a single expression peak during the second half of the light period. The new cluster B is characterized by a shift in the timing of the two peaks and a change in amplitude. A large number of genes

were detected to be expressed differently between the two strains, but are associated to clusters with similar profile. These genes are more affected on their amplitude, and not necessarily on the time regulation.

91 genes that were found in WT cluster 5, are now found in *slm1* cluster 3&4. This represents a shift in time that is reflected in an earlier expression of these genes. This correlates with the observed early expression of genes along PC2 in phase II. A substantial number of genes from WT cluster 1 are found in *slm1* cluster 2, suggesting that these genes are shifted towards later expression in time and are not expressed during the dark period anymore. Again, the transfers of genes from WT clusters 4 and 5 to *slm1* cluster B reflect an earlier change of expression. Opposite to the previously mentioned changes, which all indicate decreased expression during dark, the genes in *slm1* cluster 1&6 are now expressed exclusively during the dark period.

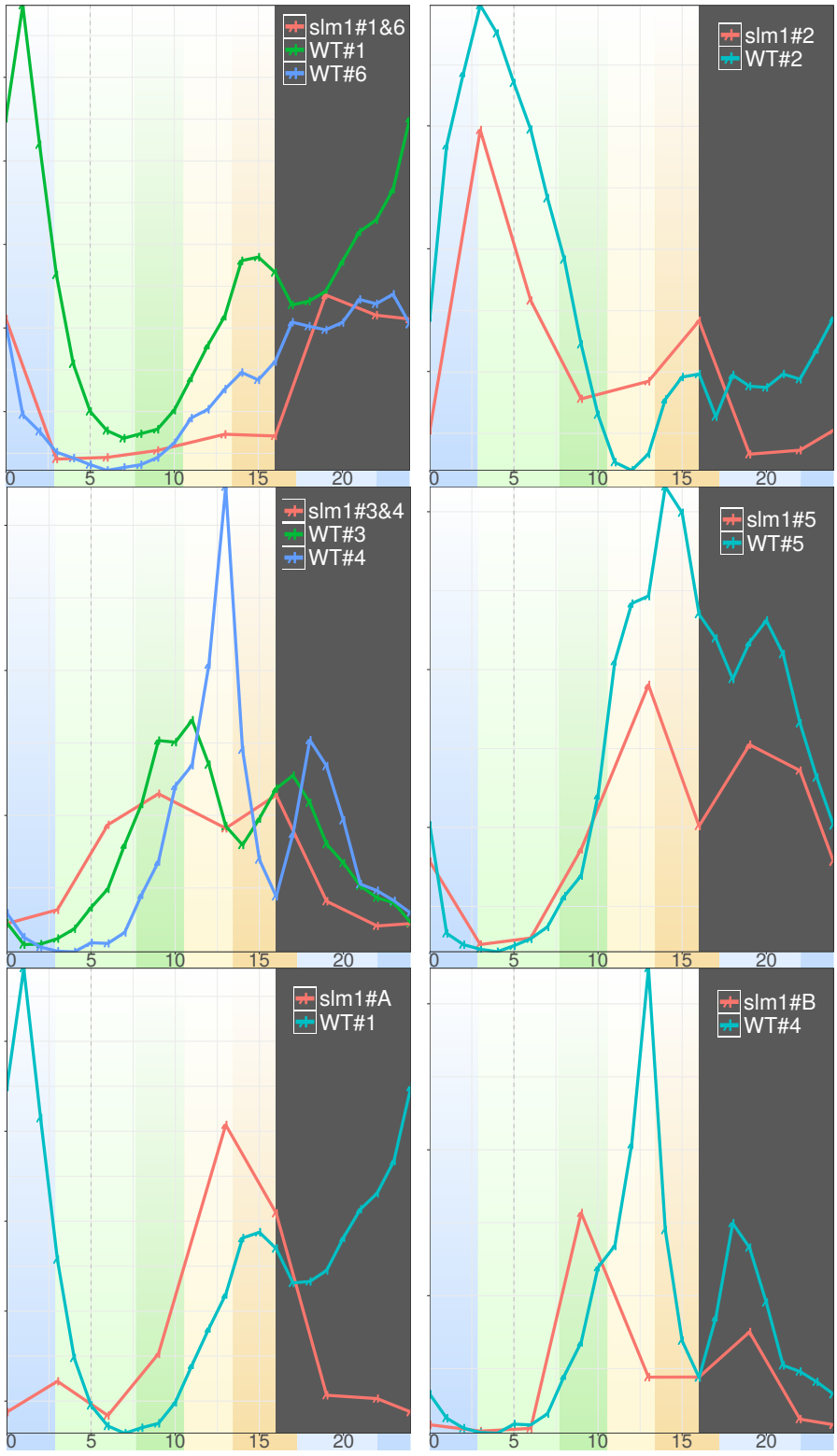


Figure 4: *Expression profiles of clusters obtained with slm1 samples and their associated profiles in the WT. WT clusters 1 and 6, slm1 cluster 1&6; WT and slm1 cluster 2; WT clusters 3 and 4, slm1 cluster 3&4; WT and slm1 cluster 5; cluster A; WT cluster 4 and slm1 cluster B. Slm1 clusters are plotted with the related WT clusters according to the common conserved genes and the redistribution of genes as in Figure 3. The background colors correspond to the colors given to the time phases in Figure 1. The gray area represents the dark period.*

3.4.3 Functional differences between gene clusters in *slm1* and WT

As observed by the similarities and changes of expression between the two strains (Figure 3B), genes have related expression in large groups. As a result, the enrichments are consistent with the large groups of genes changing expression or not. The detailed enrichments are available in supplementary file S4. Enrichments of *slm1* clusters A and B are respectively similar to WT clusters 1 and 4. *slm1* cluster 3&4 is not enriched for the two AA pathways found in WT cluster 3 (lysine degradation and phenylalanine metabolism). This cluster also remains enriched in cellular structure and DNA replication, but without any biological process GO terms. Unlike for WT cluster 6, *slm1* cluster 1&6 does not display enrichment for other fatty-acids related pathways. AA synthesis and protein translation activities are now enriched in *slm1* cluster A, but were enriched in WT cluster 1. In *slm1*, nitrogen metabolism is no longer found enriched in cluster 5. This is due to three EC numbers (out of the five) that are associated to *slm1* cluster A. Globally, the enrichments are following the expectations drawn from the gene transfers described in Figure 3B.

4 Discussion

This study analyzed gene expression dynamics in a 16:8h light-dark (LD) cycle synchronized culture of *T. obliquus* wild-type (WT). Hourly sample collection and RNA sequencing allowed us to observe in detail the changes of gene expression during a diurnal cycle. A starchless mutant of the same species (*slm1)* was also studied. *Slm1* was cultivated in the same conditions, but RNA samples were taken every three hours. The succession of cellular events were then compared between WT and *slm1*, allowing us to examine the role and importance of starch as a transient energy storage compound in this microalga.

In this work, RNA measurements were done in duplicate, where the two samples were taken from separate photobioreactors (biological duplicates). Due to restrictions on working hours of the laboratory, the samples were collected in two successive light settings in such a way that sampling could always be done during working hours. First, samples were collected on the first half of the cycle, then light settings were shifted 12

hours and the culture was then allowed to reach an oscillating steady-state before collecting samples for the second half of the cycle. The first and the last samples of each time settings are overlapping samples for control (corresponding to time points 0h and 13h). Figure 1A and C show the high similarity in expression between these samples which indicates the robustness of these cultures and the validity of the approach.

4.1 The diurnal rhythm of *T. obliquus wild-type* is driven by six transcriptional phases

4.1.1 Transcriptional phases

In this experimental setup, *T. obliquus* cells are synchronized to the diurnal light-dark (LD) cycles [8]. Many organisms synchronize their metabolism to anticipate the changes in the environment [9,48,49]. For photosynthetic microorganisms, this synchronization provides a benefit as they can capture sunlight efficiently during the day and perform light sensitive processes at night [6,12]. In T. *obliquus*, synchronization was observed in growth and cell division, as well as in changes in biomass composition [8]. The overall analysis of the changes in gene expression in both the WT and *slm1* strain shows a circular pattern in the PCA plot over two strong principal components. In WT, changes in expression appeared to occur sequentially, back and forth along each PC. In the studied process, the two main effects impacting gene expression are light availability and time itself.

By analyzing similarities between samples (Figure 1) and correlating genes expression profiles (clusters), we identified six time phases and six clusters of genes, respectively. The agreement between these two independent analyses indicates that they are related to each other. In fact, the time phase changes correspond to jumps in expression between the time points, when either a large increase or large decrease in overall gene expression occurs. The gene clusters correspond to synchronized peaks of expression. Thus, the identified phases are associated to peaks in expression for each clusters. Accordingly, cluster 1 peaks during phase Ia; cluster 2 during IIa; cluster 3 during IIb, cluster 4 during IIIa, cluster 5 during IIIb, and, finally, cluster 6

during Ib. Because the clusters regroup genes from the same biological processes, the six transcriptional phases correspond to a logical succession of biological processes.

4.1.2 Temporal succession of cellular events as described by the gene clusters

An overview of the changes in the transcriptional landscape and experimental measurements of WT through the diurnal LD cycle is presented in Figure 5.

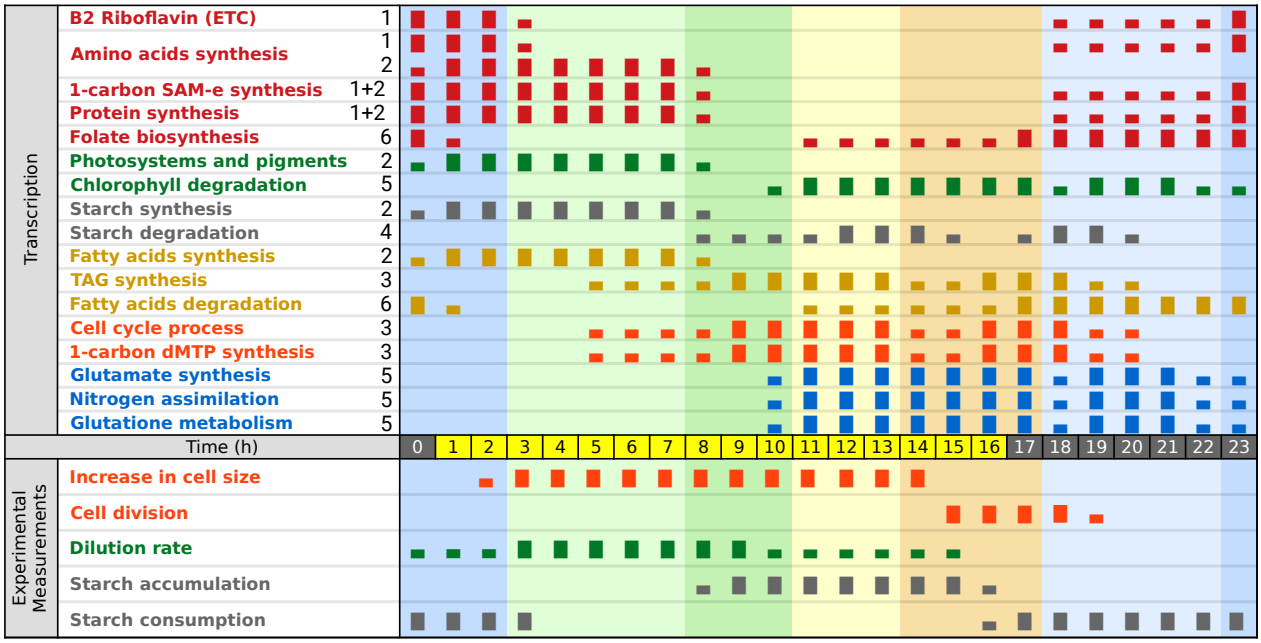


Figure 5: *Overview of the diurnal changes in the transcriptional landscape and experimental measurements of* T. obliquus *wild-type under 16:8h light/dark cycles. Experimental measurements were taken from [8]. Small and big rectangles indicate that the process is occurring at relatively low or high level, respectively. The cluster ID associated to the expression profile is indicated by the black number in the first column. Groups of related processes are indicated in the same color. The background colors correspond to the colors given to the time phases in Figure 1. Amino-acids biosynthesis two lines reflect the synthesis of different amino-acid fitting different cluster profiles.*

The expression of genes in cluster 1 starts to increase during the dark period and peaks the first after the light is switched on. The enrichment of cluster 1 reflects amino acids (AA) and protein synthesis and suggests increased ATP synthesis. These transcriptional changes point to increased catabolism for the production of energy (ATP) and agree with the observe starch consumption during the night and at the beginning of the

day. The energy and intermediates of catabolism are apparently used to synthesize amino acids for proteins synthesis as part of cell growth that starts at the start of the day.

Cluster 2 mainly contains genes involved in pigment and starch synthesis, besides carbon fixation and AA synthesis. The expression profile of this cluster matches the dilution rate profile, although gene expression precedes the dilution rate with about 3 hours [8], as can be seen in Figure 5. The expression profile of processes related to starch synthesis precedes starch accumulation by 8 hours while starch degradation at transcriptomic level appears correlated to the measured starch accumulation (Figure 5). Starch is a transient energy source and is subject to simultaneous synthesis and degradation [8,50], which results in daily fluctuations between net production during the day and consumption during the night and early day. Between 16h and 0h, degradation occurs presumably to provide energy for the night processes, since light is not available. Between 0-4h, starch is still consumed and it appears that it is used to supply additional energy and carbon for the build up of new photo systems, which is also up regulated at this time. Between 4-8h, starch synthesis increases following the up regulation of the starch synthesis genes in cluster 2. The starch synthesis machinery is getting fully operational and energy demanding processes, such as AA and protein synthesis, are down regulated, resulting in net starch synthesis.

Genes in cluster 3 show a peak in expression at 13h. GO enrichments point towards cell division which agrees with the observed cell division that starts just before the night and continues during the early night. Shortly after the genes in cluster 3 peak, genes associated to cluster 4, related to starch degradation, increase in expression. This agrees with the observed starch degradation during the night.

The next cluster for which the gene expression shows a peak is cluster 5, enriched in carbohydrates pathways, glycolysis and contains the ATP synthase. These point to starch degradation and glycolysis as the means to generate the needed energy during the night and to generate precursors, for example for glutamate metabolism. Glutamate associated genes are also found in cluster 5. Glutamate synthesis, which is at the center of nitrogen assimilation and amino-acid metabolism, seems to be upregulated before and during the night as anticipation for the synthesis of AA (cluster 1) and pigment (chlorophyll and carotenoids, cluster 2) during the night and early light period. It also suggests that *T. obliquus* continues to assimilate nitrogen

during the night, which is in agreement with the observed nitrate consumption during the night by WT (data not shown). Notably, *slm1* does not consume nitrate in the night and nitrogen metabolism is not enriched for *slm1*, which suggests that pathways in nitrogen metabolism are regulated by the energy status of the cell. Cluster 5 also contains genes associated to glutathione reductive cycle, indicating some form of oxidative stress. This stress could be a side effect of the whole day exposure to light leading to accumulation of reactive oxygen species and damaged photosystems.

Finally, cluster 6 is enriched in fatty acid degradation, which appears to occur during the night and early morning. However, fatty acid degradation was not observed experimentally. Possibly, these enzymes are involved in remodeling of membranes that has to occur after cell division, or maybe are involved in recycling unsaturated fatty-acids damaged from oxidative stress.

Folate (vitamin B9) is a vital cofactor, notably in thymidylate (dTMP) synthesis and in S-adenosylmethionine (SAM-e) synthesis. The gene expression in the pathway “one carbon pool by folate” of WT (supplementary file S5) is clearly decoupled for these two processes. While dTMP related reactions are associated to cluster 3, SAM-e reactions are associated to clusters 1 and 2. Figure 5 shows their respective synchronization with cellular division and protein synthesis. Considering the temporal succession of these events, it seems that folate is being accumulated before the light period in order to sustain the high levels of protein synthesis at dawn.

Models of a multiple fission cell cycle have been established for *C. reinhardtii* and *Scenedesmus quadricauda*where [51,52], the latter from the same family as *T. obliquus.* Comparison to the established cell cycle model, suggests that processes leading to the cell growth (G1), DNA replication (pS), mitosis (G2), and cellular division (G3), are associated to genes in clusters 2, 3, 4, and 5, respectively and correlate to phases IIa, IIb, IIIa and IIIb, respectively. In our experimental condition, the doubling time was measured to be 0.67 day in WT and 0.75 day in *slm1*, resulting in 1.5 and 1.33 divisions per day respectively. This can be explained by a double fission for a portion of the cells, 50% for WT and 33% for *slm1*. Time profiles of clusters 3, 4 and 5 show a double peak of expression 4 to 5 hours apart. This second and lower peak could be associated to that portion of cells performing a double fission.

In summary, the annotation and enrichment analyses uncovered the general function of the genes in each cluster, thus showing a clear succession of cellular events: RNA transcription, gene expression, protein synthesis, pigments synthesis, starch synthesis, fatty acids synthesis, protein glycosylation, cellular growth and division, starch degradation and sugars inter-conversions, riboflavin synthesis, folate synthesis, and, finally, fatty acids degradation (Figure 5). This succession indicates that *T. obliquus* does not simply adapt its cellular phenotype to direct signals from the light switch, but it also anticipates the changes in light conditions.

4.2 Impact of starch deficiency on the diurnal rhythm

When analyzing data from *slm1*, significant changes in the temporal profiles were identified for the vast majority of the genes compared to the WT (3902 genes). In most cases, the genes remained clustered together, as for the WT, and most differences can be explained by a time shift or a change in amplitude. Time changes of expression in phase II (and to a lesser extent IIIa) contains most of the overall differences between WT and *slm1*. However, this could be due to the less frequent sampling of *slm1* (every 3h). From the six clusters of slm1, cluster A showed a really novel expression profile not seen in WT. In WT the profile of these genes more or less matches the observed net consumption of starch that occurs during the night and early day. In other words, impossibility to consume starch during this period in *slm1* may be related to the low expression of the genes occurring in cluster A.

The biochemical data for *slm1* also match the transcriptome results, but to a lesser extent than for WT. Data obtained from the starchless mutant *slm1* shows that the phases from the WT are to some extent preserved, and starch deficiency seems to result in rescheduling biological processes from the dark to the light period. It is possible that these processes could not happen at night due to the lack of energy originally provided by starch degradation. Processes required for expression and translation were transferred from cluster 1 to cluster A in *slm1*. Thus, expression of these genes seems to depend on the presence of starch and is possibly

related to the energy status of the cell. In *slm1*, the expression of these processes is shifted to the time when starch accumulates in the WT during which it may be assumed that sufficient energy is available.

Genes in *slm1* cluster 1&6 are mainly expressed during the dark period. This indicates that the scheduling of processes that occur in the dark period and that do not depend on energy sources only suffers small changes. Notably, genes associated to fatty acid degradation in *slm1*, are in this cluster, although no significant reduction in TAG content was observed.

The reason of the incapacity of *slm1* to synthesize starch has already been studied [17]. The gene coding for the small subunit of ADP-glucose pyrophosphorylase was found to contain a non-sense mutation. However *slm1* has been obtained with UV, that is known to produce multiple mutations in genomes, therefore we can not rule out differences in gene regulation due to other mutation not related to starch deficiency.

4.3 Selected reactions and pathways

In addition to the described time shifts caused by starch deficiency, a number of processes show differences in expression between *sml1* and WT that seem more complex. These are discussed in the following paragraphs.

4.3.1 Starch synthesis

The diurnal expression of genes related to the reactions analyzed in this section are depicted in Figure 6. Three of the four enzymes necessary for starch synthesis are found in cluster 2, being the ADP-Glucose pyrophosphorylase (EC:2.7.7.27), the starch synthase (EC:2.4.1.21) and the granule-bound starch synthase (EC:2.4.1.242). The fourth, starch branching enzyme, was not found to be regulated in time. Genes associated to the starch synthase (g7327.t3 and g2865.t1), show complementary profiles in WT, each being highly expressed during one half of the light period. However, their expression appears stretched during the light period in *slm1*. This could be the response to the lack of ADP glucose. Similarly the granule-bound starch synthase expression peaks in the middle of the light period for the WT, but appears stretched over the light period in *slm1*. The difference in time regulation of the two types of starch synthase could be explained

by their associated processes on starch branching nature [53]. The amylose isomerase (EC:2.4.1.18) was associated to two genes, with very similar profiles, but without apparent time regulation for either strain.

The reason of the incapacity of *slm1* to synthesize starch has already been studied [17].The gene coding for the small subunit of ADP-glucose pyrophosphorylase was found to contain a non-sense mutation. In conditions of continuous light, the gene coding for the small subunit of ADP-glucose pyrophosphorylase (known to be mutated in *slm1)* was found to be strongly down-regulated in comparison to WT (approx. 5 folds). In this study, we found that this gene (g788.t1) is even more strongly down-regulated, meaning that the changes become even more prominent in presence of a diurnal cycle. Additionally, we found that the time regulation is preserved between the two strains.

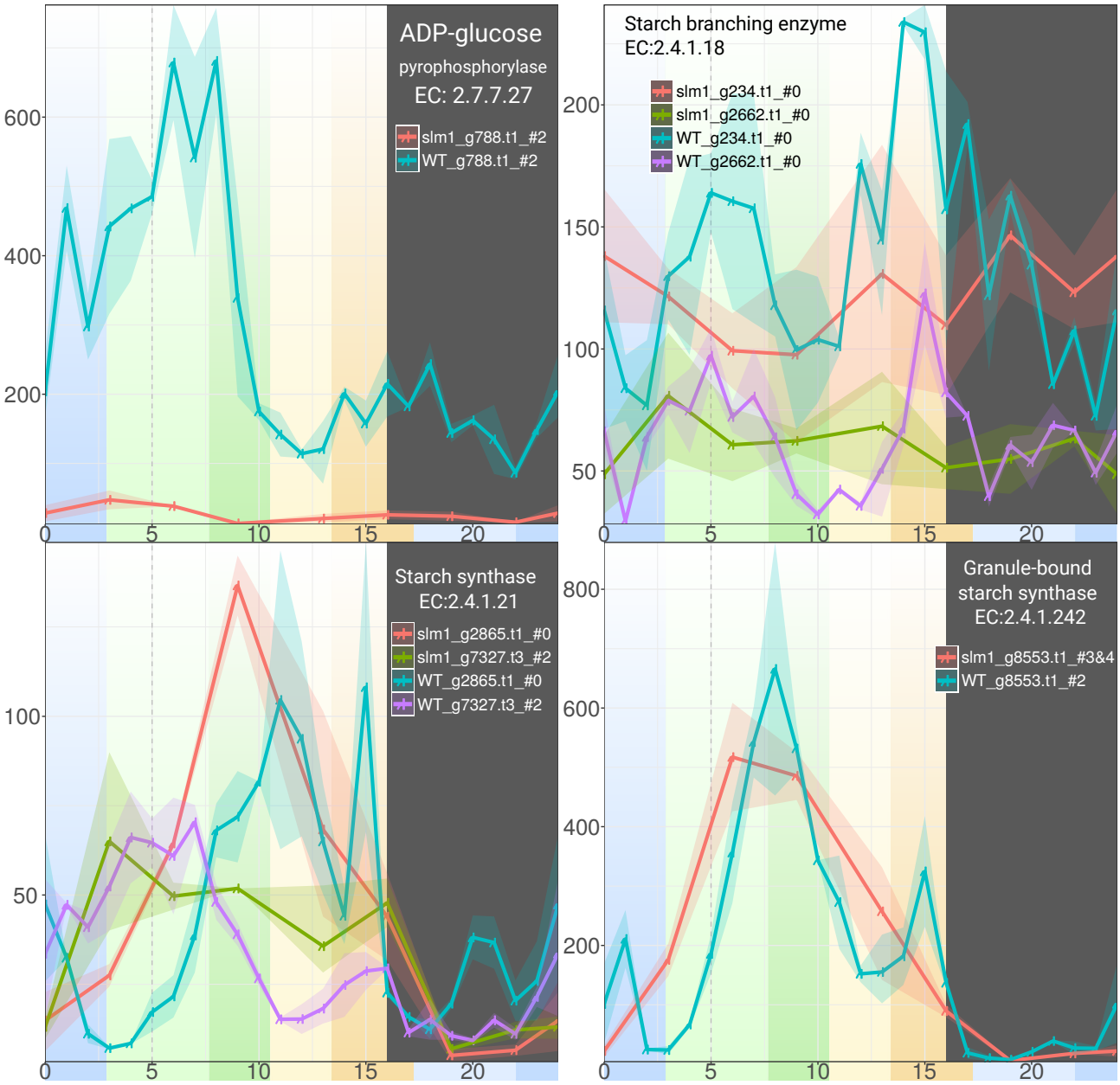


Figure 6: *Diurnal expression of genes related to carbon fixation. The points represent the mean for each time points, the ribbon covers the minimum and maximum values for each time points. The genes associated to the same EC number are plotted together. All genes in the same plot are using the same y axis scale (not shown), but the scale between each of the six plots varies. Each plot is labeled with the reaction name, the corresponding EC number, and the color legend for each gene-strain combinations. For each gene-strain combination, the letter or the number behind the # symbol, indicate the cluster with which the gene is associated. The background color corresponds to the time phases from Figure 1. The dark gray area corresponds to the dark period.*

4.3.2 Carbon fixation

The diurnal expression of the genes related to the reactions analyzed in this section is depicted in Figure 7. The expression of RuBisCO (EC:4.1.1.39) is noisy, although it is generally lower in *slm1* than in WT. The logical explanation is that less carbon can be fixed due to starch deficiency. Similarly, the bisphosphatase

profile is noisy but with a clear pattern of higher expression between 4h and 9h in WT that is not observed in *slm1*.Other reactions associated to carbon fixation are fitting the expression pattern of clusters 1 or 2 in both WT and *slm1*. Among these, the two enzymes preceding in the carbon fixation pathway (phosphoriboisomerase and phosphopentokinase) follow an expected pattern with carbon fixation during the light with a strong peak at the beginning of the light period.

The RuBisCO activase is activated by light and helps to release the substrate from the RuBisCO active site, accelerating the reaction [54,55]. Three candidate genes were identified among which only g6221.t1 displays high expression and a detected time profile, shown in Figure 7. Overall, the expression of the RuBisCO activase resembles the expression of the RuBisCO, but with an higher amplitude and a few hours earlier.

In complement to the RuBisCO activase, the RuBisCO inhibitor 2-Carboxy-D-arabitinol 1-phosphate (CA1P) is inactivated by the 2-carboxy-D-arabinitol-1-phosphatase (CA1Pase EC: 3.1.3.63). CA1P normally binds to RuBisCO under dark conditions, which prevents it from performing chemical reactions. In WT the gene g18093.t1, associated to the CA1Pase, displays a rather constant expression through both light and dark periods, except for a dip in expression around 1-2 h. However, in *slm1*, CA1Pase has a higher expression during the dark period and lower during the day.

All the aforementioned enzymes are co-localized in the pyrenoid with the RuBisCO [56]. This sub-cellular micro-compartment plays a major role in carbon fixation and it is logically localized in the chloroplast, along the thylakoid membrane. It is found in green algae (chlorophyta), and therefore found in *T. obliquus* [57]. Pyrenoids are not delimited by a membrane, but they accumulate starch at their periphery in the form of a sheath. It is believed that starch is a barrier that limits the transfer of CO_2 to the rest of the chloroplast and outer compartments. It was demonstrated that the pyrenoid plays a role in the CCM, allowing higher levels of carbon fixation in *C. reinhardtii* [58], however analysis of a *C. reinhardtii* starchless mutant showed that the starch sheath surrounding the pyrenoid is not involved in the CCM [59]. Therefore, the starchless mutant, *slm1*, should not suffer from a lessened CCM in comparison to the wild-type. The lower yield on light should be due to the lack of a transient energy storage allowing for harvesting more light energy during the day and using this in the night.

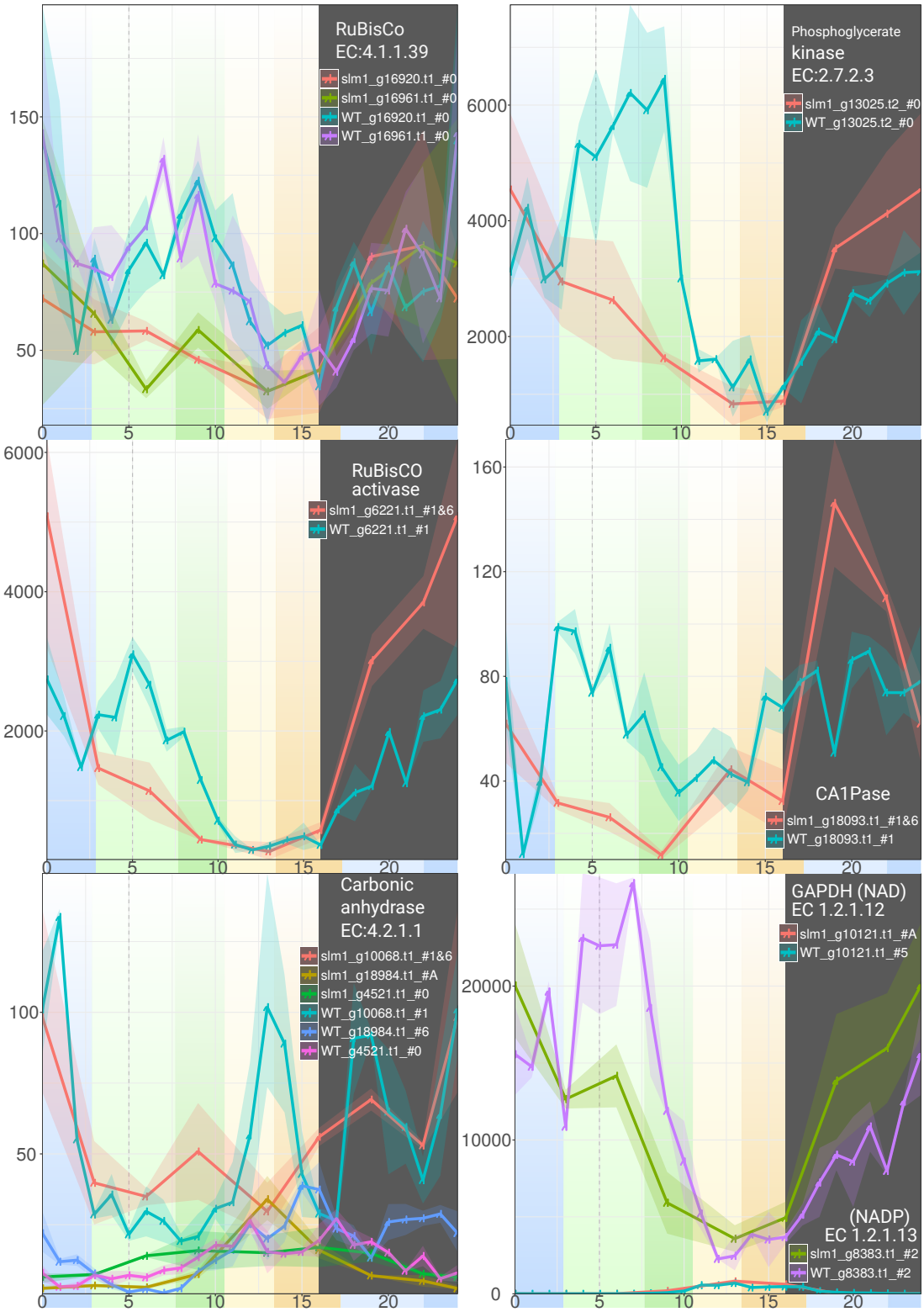


Figure 7:*Diurnal expression of genes related to carbon fixation. The points represent the mean for each time points, the ribbon covers the minimum and maximum values for each time points. The genes associated to the same EC number are plotted together. All genes in the same plot are using the same y axis scale (not shown), but the scale between each of the six plots vary. Each plot is labeled with the reaction name, the corresponding EC number, and the color legend for each strain_gene_cluster combination. For each gene-strain combination, the letter or the number behind the # symbol, indicates the cluster in which the gene was found. The background color corresponds to the time phases from Figure 1. The dark gray area corresponds to the dark period.*

4.3.3 Nitrogen metabolism

The diurnal expression of the genes related to the reactions analyzed in this section are depicted in the Figure 8. The first enzymes of the nitrogen assimilation pathways are nitrate and nitrite reductases converting nitrate to ammonium. These enzymes are known to be strongly regulated by the light and nitrogen condition of the cell [60]. We were not able to identify the nitrate reductase gene in the transcriptome. The nitrite reductase gene displayed extremely low expression and no identifiable diurnal profile. Ammonium can be transformed into glutamate via two pathways, either by the glutamate dehydrogenase (GDH, EC 1.4.1.3 and 1.4.1.4), or via the glutamate synthase cycle that consists of glutamine synthase (GS, EC 6.3.1.2) followed by glutamate synthase (EC 1.4.1.14 and 1.4.7.1). GS is known as the main contributor for nitrogen assimilation [61], whereas GDH is generally involved in balancing nitrogen between metabolites by recycling ammonium for GS.

The candidate gene for GS (g10416.t1)was located in the chloroplast [61, 62]. and showed higher expression at night and early day with similar profiles in WT and *slm1*. Expression during the dark period seems to peak at the same time as the GDH (g547.t1). For the glutamate synthase, we have three candidates, two using NAD as a cofactor, and one using ferredoxin as a cofactor. The two transcripts associated to the NAD-dependent reaction peak at the end of the day and have a similar profile in WT and *slm1*. WT expression of the ferredoxin-dependent reaction is very similar to GS but with a small time-shift (approximately 2h) towards later expression. Additionally, we also notice a small peak of expression right before the dark period that comes right after the peak of the two genes associated to the NAD-dependent glutamate synthase. Additionally, the ferredoxin-dependent glutamine synthase also peaks right after the GDH during the late light period, but also several hours after the GDH peak during the dark period. These observations in WT match the observed behavior of these enzymes in *C. reinhardtii* [63], meaning that the ferredoxin-dependent enzyme cannot only perform the assimilation during the light period, but also contributes to the assimilation during the dark period. Expression of the ferredoxin-dependent enzymes in *slm1* follows a similar temporal profile but is strongly reduced.

Two genes were associated to GDH EC 1.4.1.3, namely g547.t1 and g12033.t1.The second one, g12033.t1, was also associated to EC 1.4.1.4, although with a lower score. Both genes have similar WT expression profiles. In *slm1*, g547.t1 exhibits a similar profile whereas g12033.t1 in *slm1* appears to have lost sharp peaks with very gradual variation. The opposing expression of GDH and GS matches previous observations in *C. reinhardtii* [63], suggesting GDH to be an exclusively catabolic enzyme. These profiles suggest nitrogen recycling at the end of the light period and beginning of dark. Differences in the GDH are probably related to differences in cofactor utilization [60]. Nitrogen concentration measurements (data not shown) revealed that *slm1* does not assimilate nitrogen during the dark period, which would suggest EC 1.4.1.4 as the main route for nitrogen fixation during the dark period.

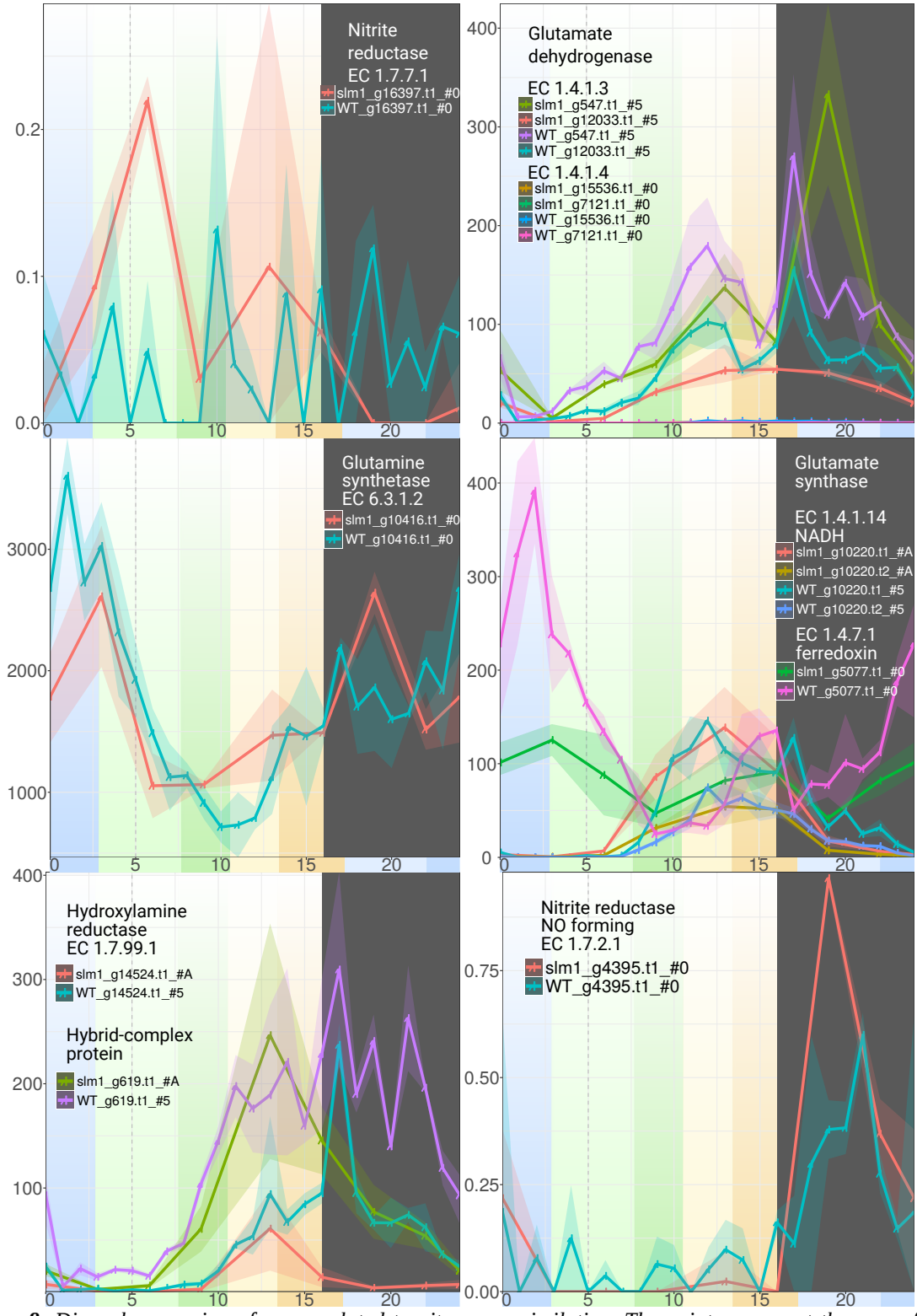


Figure 8: *Diurnal expression of genes related to nitrogen assimilation. The points represent the mean for each time points, the ribbon covers the minimum and maximum values for each time points. The genes associated to the same EC number are plotted together. All genes in the same plot are using the same y axis scale (not shown), but the scale between each of the five plots vary. Each plots are labeled with the reaction name, the corresponding EC number, and the color legend for each strain_gene_cluster combinations. For each gene-strain combination, the letter or the number behind the # symbol, indicate the cluster in which the gene was found. The background color corresponds to the time phases from Figure 1. The gray area corresponds to the dark period.*

5 Conclusions

Oleaginous microalgae are a promising source of biofuels. Among the oleaginous microalgae, *Tetradesmus obliquus* is a promising candidate that can reach a high maximum TAG yield on light and TAG content under nitrogen starvation. However, in order to develop strategies to enhance lipid productivity for commercial production, a system-level understanding of metabolism is essential. Large scale microalgal production will be done outdoors under natural LD cycles. Therefore, the impact of LD cycles has to be carefully considered when characterizing the behavior of microalgae.

In this work, we show that LD cycles induce systems level transcriptional changes in 4686 genes showing a clear expression pattern and indicating a strict succession of cellular events. These cellular events were found to be in accordance with the biochemical measurements. While some regulations seem a direct response to light, other regulations reflect an anticipation to the switch from light to dark and vice-versa. This observed anticipation indicates an inner time-keeping system.

Additionally, we studied the diurnal transcriptional changes in the starchless mutant of *T. obliquus slm1*. Pathways directly associated to energy storage, such as carbon fixation, and other processes such as nitrogen metabolism are strongly affected in *slm1*. The large changes in activity of these processes are attributed to starch deficiency, which seems to be the only transient energy storage compound in *T. obliquus*. Genes associated to TAG and lipid degradation are highly expressed during the dark period in *slm1*. This suggests TAG as a transient energy storage, however, no significant changes in TAG content were measured. As a result of the lack of an energy source during the dark, some of the energy related processes shift to the light period and the algae are less well prepared to start the next cycle at the beginning of the day. More subtly, processes related to the cellular cycle were detected to start earlier in *slm1* than in WT.

Overall, we provide for the first time a diurnal transcriptional landscape under LD cycles with a high resolution (1h intervals) of an oleaginous green microalgae that produces both starch and TAG under nitrogen starvation. This is also the first time a diurnal transcriptional landscape is described and compared for a starchless mutant of green algae. The diverse set of insights revealed in this analysis of the diurnal cycle

of *T. obliquus* is very valuable to develop strategies to increase yields. We suggest that the presented transcriptional landscape should be carefully considered when designing future experiments with LD cycles, including metabolic engineering approaches.

Funding

This research project was supported by the Consejo Nacional de Ciencia y Tecnología – CONACYT, Mexico, Scholar 218586/Scholarship 314173. In addition, GMLS is part of the program “Doctores Jóvenes para el Desarrollo Estratégico Institucional” by the Universidad Autónoma de Sinaloa. This work has been financially supported by the Systems Biology Investment Program of Wageningen University, (KB-17-003.02-29).

Acknowledgments

The authors would like to thank Tom Schonewille for the RNA extraction of all samples.

Author Contributions

BMC, GMLS, PJS, DV, VAPMS, RW and MSD conceived the research and designed the experiments; GMLS and IMR performed the experiments. BMC analyzed the data. BMC, GMLS, PJS, DEM and MSD interpreted the data. BMC and GMLS wrote the article. PJS, DEM and MSD supervised and edited the manuscript. IMR, DV, VAPMS and RW critically revised the manuscript. VAPMS and RW obtained funding. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Supplementary Files

S1: Annotation. The association between gene, protein, EC number, GO terms is summarized in this table sheet. The protein sequences are also given in this file.

S2: PCA and heatmap of all expressed genes, before time profile selection. In the PCA of WT, the two first component explain a big majority of the variations, and the time points follow a clear ordered circular pattern. The PCA with both strains displays a more noisy pattern and the

S3: Finding optimal hierarchical cluster separation. Two methods were used to find the optimal separation of genes into clusters. The first figure contains a plot for each index calculated for every cluster separation steps. The second figure contains the heatmap representation of the Rand index values generated by comparing cluster separation one on one. All these indexes were generated from R package “cluster.stats”.

S4: Cluster enrichment. Results of the enrichment analyses for each clusters of each strains, in both pathways and GO terms. Separate enrichment are also displayed for the gene with conserved expression between strains and those with altered expression, described in table 1 and Figure 3.

S5: Images of cluster colored KEGG pathway maps. Compressed archive file of every annotated maps in PNG format. EC numbers colored according to related genes cluster association. Each EC number box is divided in seven parts: the first for genes with no time regulation and therefore no cluster association, the following six to the clusters defined in the strain. The colors and positions are depicted in the legend file called “_0_color_legend.png”. For WT, the colors are the same as in Figure 2. For *slm1*, the colors are given based on their WT counterpart when possible. Gene not regulated in time, filtered out by our analysis and by extend not clustered were added on the maps as gray color.

6 References

[1] M. T. Guarnieri and P. T. Pienkos, “Algal omics: unlocking bioproduct diversity in algae cell factories,” *Photosynthesis research*, vol. 123, no. 3. pp. 255–263, 2015.

[2] R. H. Wijffels and M. J. Barbosa, “An outlook on microalgal biofuels,” *Science*, vol. 329, no. 5993. pp. 796–799, 2010.
[3] R. H. Wijffels, M. J. Barbosa, and M. H. M. Eppink, “Microalgae for the production of bulk chemicals and biofuels,” *Biofuels, Bioproducts and Biorefining*, vol. 4, no. 3. pp. 287–295, 2010.
[4] W. Blanken, M. Cuaresma, R. H. Wijffels, and M. Janssen, “Cultivation of microalgae on artificial light comes at a cost,” *Algal Research*, vol. 2, no. 4. pp. 333–340, 2013.
[5] N. H. Norsker, M. J. Barbosa, M. H. Vermuë, and R. H. Wijffels, “Microalgal production - A close look at the economics,” *Biotechnology Advances*, vol. 29, no. 1. pp. 24–27, 2011.
[6] L. De Winter, A. J. Klok, M. Cuaresma, M. J. Barbosa, and R. H. Wijffels, “The synchronized cell cycle of Neochloris oleoabundans and its in fl uence on biomass composition under constant light conditions,” *Algal Res.*, vol. 2, no. 4, pp. 313–320, 2013.
[7] E. M. Farré, “The regulation of plant growth by the circadian clock,” *Plant Biology*, vol. 14, no. 3. pp. 401–410, 2012.
[8] G. M. León-Saiki, I. M. Remmers, D. E. Martens, P. P. Lamers, R. H. Wijffels, and D. van der Veen, “The role of starch as transient energy buffer in synchronized microalgal growth in Acutodesmus obliquus,” *Algal Res.*, vol. 25, pp. 160–167, 2017.
[9] C. R. McClung, “Plant Circadian Rhythms,” *PLANT CELL ONLINE*, vol. 18, no. 4, pp. 792–803, 2006.
[10] G. M. León-Saiki, T. C. Mart\’\i, D. van der Veen, R. H. Wijffels, and D. E. Martens, “The impact of day length on cell division and efficiency of light use in a starchless mutant of Tetradesmus obliquus,” *Algal Res.*, vol. 31, pp. 387–394, 2018.
[11] S. S. Nikaido and C. H. Johnson, “Daily and circadian variation in survival from ultraviolet radiation in Chlamydomonas reinhardtii.,” *Photochem. Photobiol.*, vol. 71, no. 6, pp. 758–765, 2000.
[12] L. Suzuki and C. H. Johnson, “Algae know the time of day: Circadian and photoperiodic programs,” *Journal of Phycology*, vol. 37, no. 6. pp. 933–942, 2001.
[13] G. Breuer, P. P. Lamers, D. E. Martens, R. B. Draaisma, and R. H. Wijffels, “The impact of nitrogen starvation on the dynamics of triacylglycerol accumulation in nine microalgae strains,” *Bioresour. Technol.*, vol. 124, pp. 217–226, 2012.
[14] I. M. Remmers, A. Hidalgo-Ulloa, B. P. Brandt, W. A. C. Evers, R. H. Wijffels, and P. P. Lamers, “Continuous versus batch production of lipids in the microalgae Acutodesmus obliquus,” *Bioresour. Technol.*, vol. 244, pp. 1384–1392, 2017.
[15] G. Breuer, P. P. Lamers, D. E. Martens, R. B. Draaisma, and R. H. Wijffels, “Effect of light intensity, pH, and temperature on triacylglycerol (TAG) accumulation induced by nitrogen starvation in Scenedesmus obliquus,” *Bioresour. Technol.*, vol. 143, no. 0, pp. 1–9, 2013.
[16] L. De Jaeger *et al.*, “Superior triacylglycerol (TAG) accumulation in starchless mutants of Scenedesmus obliquus: (I) mutant generation and characterization,” *Biotechnol. Biofuels*, vol. 7, no. 1, p. 69, 2014.
[17] L. de Jaeger, *Strain improvement of oleaginous microalgae*. Wageningen University, 2015.
[18] S. M. Smith *et al.*, “Diurnal Changes in the Transcriptome Encoding Enzymes of Starch Metabolism Provide Evidence for Both Transcriptional and Posttranscriptional Regulation of Starch Metabolism in Arabidopsis Leaves 1,” *Plant Physiol.*, 2004.
[19] S. R. Smith *et al.*, “Transcriptional Orchestration of the Global Cellular Response of a Model Pennate Diatom to Diel Light Cycling under Iron Limitation,” *PLoS Genet.*, 2016.
[20] M. S. Chauton, P. Winge, T. Brembu, O. Vadstein, and A. M. Bones, “Gene Regulation of Carbon Fixation, Storage, and Utilization in the Diatom Phaeodactylum tricornutum

Acclimated to Light/Dark Cycles1[C][W][OA].," *Plant Physiol.*, vol. 161, no. 2, pp. 1034–1048, 2013.
[21] E. Poliner *et al.*, "Transcriptional coordination of physiological responses in Nannochloropsis oceanica CCMP1779 under light/dark cycles," *Plant J.*, vol. 83, no. 6, pp. 1097–1113, 2015.
[22] J. M. Zones, I. K. Blaby, S. S. Merchant, and J. G. Umen, *High-Resolution Profiling of a Synchronized Diurnal Transcriptome from Chlamydomonas reinhardtii Reveals Continuous Cell and Metabolic Differentiation*, vol. 27, no. October. 2015.
[23] S. Suzuki, K. I. Ishida, and Y. Hirakawa, "Diurnal transcriptional regulation of endosymbiotically derived genes in the chlorarachniophyte bigelowiella natans," *Genome Biol. Evol.*, vol. 8, no. 9, pp. 2672–2682, 2016.
[24] T. Fujiwara *et al.*, "Periodic gene expression patterns during the highly synchronized cell nucleus and organelle division cycles in the unicellular red alga cyanidioschyzon merolae," *DNA Res.*, vol. 16, no. 1, pp. 59–72, 2009.
[25] M. J. Wynne and J. K. Hallan, "Reinstatement of Tetradesmus G. M. Smith (Sphaeropleales, Chlorophyta)," *Feddes Repert.*, 2016.
[26] B. M. Carreres *et al.*, "Draft Genome Sequence of the Oleaginous Green Alga *Tetradesmus obliquus* UTEX 393," *Genome Announc.*, vol. 5, no. 3, pp. e01449-16, 2017.
[27] K. J. Hoff, S. Lange, A. Lomsadze, M. Borodovsky, and M. Stanke, "BRAKER1: Unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS," *Bioinformatics*, vol. 32, no. 5, pp. 767–769, 2015.
[28] J. J. Koehorst, J. C. J. van Dam, E. Saccenti, V. A. P. Martins dos Santos, M. Suarez-Diez, and P. J. Schaap, "SAPP: functional genome annotation and analysis through a semantic framework using FAIR principles," *Bioinformatics*, 2017.
[29] J. C. J. Van Dam, J. J. Koehorst, J. O. Vik, P. J. Schaap, and M. Suarez-diez, "Interoperable genome annotation with GBOL, an extendable infrastructure for functional data mining," *bioRxiv*, no. 1, pp. 1–9, 2017.
[30] M. D. Wilkinson *et al.*, "The FAIR Guiding Principles for scientific data management and stewardship," *Sci. Data*, vol. 3, p. 160018, 2016.
[31] A. Dobin *et al.*, "STAR: Ultrafast universal RNA-seq aligner," *Bioinformatics*, vol. 29, no. 1, pp. 15–21, 2013.
[32] C. Trapnell *et al.*, "Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation," *Nat. Biotechnol.*, vol. 28, no. 5, pp. 511–515, 2010.
[33] P. Jones *et al.*, "InterProScan 5: Genome-scale protein function classification," *Bioinformatics*, vol. 30, no. 9, pp. 1236–1240, 2014.
[34] M. Falda *et al.*, "Argot2: A large scale function prediction tool relying on semantic similarity of weighted Gene Ontology terms," *BMC Bioinformatics*, vol. 13, no. SUPPL.4, 2012.
[35] N.-N. Nguyen, S. Srihari, H. W. Leong, and K.-F. Chong, "EnzDP: Improved enzyme annotation for metabolic network reconstruction based on domain composition profiles," *J. Bioinform. Comput. Biol.*, vol. 13, no. 05, p. 1543003, 2015.
[36] M. J. Nueda, S. Tarazona, and A. Conesa, "Next maSigPro: Updating maSigPro bioconductor package for RNA-seq time series," *Bioinformatics*, vol. 30, no. 18, pp. 2598–2602, 2014.
[37] M. Robinson and A. Oshlack, "A scaling normalization method for differential expression analysis of RNA-seq data," *Genome Biol.*, vol. 11, no. 3, p. R25, 2010.
[38] M. D. Robinson, D. J. McCarthy, and G. K. Smyth, "edgeR: A Bioconductor package for differential expression analysis of digital gene expression data," *Bioinformatics*, vol. 26, no. 1, pp. 139–140, 2009.

[39] T. Äijö *et al.*, “Methods for time series analysis of RNA-seq data with application to human Th17 cell differentiation,” *Bioinformatics*, vol. 30, no. 12, 2014.
[40] C. Hennig, “fpc: Flexible Procedures for Clustering.” 2018.
[41] R Core Team, “R: A Language and Environment for Statistical Computing.” Vienna, Austria, 2017.
[42] M. Kanehisa and S. Goto, “KEGG: kyoto encyclopedia of genes and genomes.,” *Nucleic Acids Res.*, vol. 28, no. 1, pp. 27–30, Jan. 2000.
[43] M. Kanehisa, Y. Sato, M. Kawashima, M. Furumichi, and M. Tanabe, “KEGG as a reference resource for gene and protein annotation,” *Nucleic Acids Res.*, vol. 44, no. D1, pp. D457–D462, Jan. 2016.
[44] M. Carlson, S. Falcon, H. Pages, and N. Li, “A set of annotation maps describing the entire Gene Ontology.” 2007.
[45] B. Christ *et al.*, “Cytochrome P450 CYP89A9 is involved in the formation of major chlorophyll catabolites during leaf senescence in Arabidopsis.,” *Plant Cell*, vol. 25, no. 5, pp. 1868–80, May 2013.
[46] C. Oda-Yamamizo, N. Mitsuda, S. Sakamoto, D. Ogawa, M. Ohme-Takagi, and A. Ohmiya, “The NAC transcription factor ANAC046 is a positive regulator of chlorophyll degradation and senescence in Arabidopsis leaves,” *Sci. Rep.*, vol. 6, no. 1, p. 23609, Jul. 2016.
[47] B. Christ and S. Hörtensteiner, “Mechanism and Significance of Chlorophyll Breakdown,” *J. Plant Growth Regul.*, vol. 33, no. 1, pp. 4–20, Mar. 2014.
[48] M. Mittag, “Circadian rhythms in microalgae,” *International Review of Cytology*, vol. 206. pp. 213–247, 2001.
[49] H. C. Causton, K. A. Feeney, C. A. Ziegler, and J. S. O’Neill, “Metabolic cycles in yeast share features conserved among circadian rhythms,” *Curr. Biol.*, vol. 25, no. 8, pp. 1056–1062, 2015.
[50] M. Stitt and S. C. Zeeman, “Starch turnover: Pathways, regulation and role in growth,” *Current Opinion in Plant Biology*, vol. 15, no. 3. pp. 282–292, 2012.
[51] K. Bišová and V. Zachleder, “Cell-cycle regulation in green algae dividing by multiple fission,” *J. Exp. Bot.*, vol. 65, no. 10, pp. 2585–2602, Jun. 2014.
[52] V. Zachleder, O. Schläfli, and A. Boschetti, “Growth-controlled oscillation in activity of histone H1 kinase during the cell cycle of Chlamydomonas reinhardtii (chlorophyta),” *J. Phycol.*, vol. 33, no. 4, pp. 673–681, Aug. 1997.
[53] A. Izumo *et al.*, “Effects of granule-bound starch synthase I-defective mutation on the morphology and structure of pyrenoidal starch in Chlamydomonas,” *Plant Sci.*, 2011.
[54] S.-H. Jin, D.-A. Jiang, X.-Q. Li, and J.-W. Sun, “Characteristics of photosynthesis in rice plants transformed with an antisense Rubisco activase gene.,” *J. Zhejiang Univ. Sci.*, vol. 5, no. 8, pp. 897–9, Aug. 2004.
[55] A. R. Portis and Jr., “Rubisco activase – Rubisco’s catalytic chaperone,” *Photosynth. Res.*, vol. 75, no. 1, pp. 11–27, 2003.
[56] a Lopez-Ruiz, J. P. Verbelen, J. M. Roldan, and J. Diez, “Nitrate reductase of green algae is located in the pyrenoid.,” *Plant Physiol.*, 1985.
[57] A. E. M. M. R. Afify, G. S. El Baroty, F. K. El Baz, H. H. Abd El Baky, and S. A. Murad, “Scenedesmus obliquus: Antioxidant and antiviral activity of proteins hydrolyzed by three enzymes,” *Journal of Genetic Engineering and Biotechnology*, 2018.
[58] O. D. Caspari *et al.*, “Pyrenoid loss in Chlamydomonas reinhardtii causes limitations in CO2 supply, but not thylakoid operating efficiency,” *J. Exp. Bot.*, 2017.

[59] A. Villarejo, F. Martinez, M. Pino Plumed, and Z. Ramazanov, “The induction of the CO2 concentrating mechanism in a starch-less mutant of Chlamydomonas reinhardtii,” *Physiol. Plant.*, vol. 98, no. 4, pp. 798–802, Dec. 1996.
[60] J. A. Hellebust and I. Ahmad, “Regulation of Nitrogen Assimilation in Green Microalgae,” *Biol. Oceanogr.*, 1989.
[61] N. Mallick, L. C. Rai, F. H. Mohn, and C. J. Soeder, “Studies on nitric oxide (NO) formation by the green alga Scenedesmus obliquus and the diazotrophic cyanobacterium Anabaena Doliolum,” *Chemosphere*, vol. 39, no. 10, pp. 1601–1610, Oct. 1999.
[62] M. Tardif *et al.*, “Predalgo: A new subcellular localization prediction tool dedicated to green algae,” *Mol. Biol. Evol.*, vol. 29, no. 12, pp. 3625–39, Dec. 2012.
[63] J. V. Cullimore and A. P. Sims, “Pathway of ammonia assimilation in illuminated and darkened Chlamydomonas reinhardii,” *Phytochemistry*, 1981.