Title

Surprisal analysis of the transcriptomic response of the green microalga *Chlamydomonas* to the addition of acetate during day/night cycles

Willamme R[1], Bogaert KA[2], Remacle F[2], Remacle C[1*]

[1]Genetics and Physiology of Microalgae, UR InBios, Chemin de la Vallée, 4, University of Liège, 4000 Liège, Belgium

[2]Theoretical Physical Chemistry, UR MOLSYS, Allée du 6 Août, 11, University of Liège, 4000 Liège, Belgium

Willamme R, Remi.Willamme@doct.uliege.be; Bogaert KA, bogaert.kenny@gmail.com; Remacle F, fremacle@uliege.be; Remacle C, c.remacle@ulg.ac.be

* Corresponding author: Claire Remacle, Genetics and Physiology of Microalgae, UR InBios, Chemin de la Vallée, 4, University of Liège, 4000 Liège, Belgium, email: c.remacle@ulg.ac.be

Abstract

Our study aims to find gene pathways that depend on acetate assimilation under diurnal conditions in the microalga Chlamydomonas. We compare the transcriptome of two strains, one control and one mutant deficient for the glyoxylate cycle essential in acetate metabolism, cultivated under day/night cycles with acetate. We apply surprisal analysis, an information-theoretic approach, to the RNA-seq data. Carrying out the analysis on groups of dark and light phase samples separately allows identifying constraints and gene pathways that discriminate between mutant and control samples. Carbon metabolism is the most important in the light phase for the control strain while the dark phase is enriched in cell division pathways. The mutant phenotype includes genes pathways of stress response and autophagy in the two phases. Cell division pathways are found in the light phase and catabolic pathways in the dark phase, highlighting a rewiring of the mutant transcriptome in these cyclic cultivation conditions.

Keywords

Surprisal analysis; transcriptomics; day/night cycles; isocitrate lyase; Chlamydomonas

Highlights (85 characters including spaces)

- Microalgal control and mutant unable to use acetate were compared by transcriptomics.
- Surprisal analysis was applied to RNA-Seq data sampled during a day/night cycle.
- Constraints of samples grouped on the day/night phases allowed strain separation.
- Strain-specific gene pathways were identified in the two phases.

Introduction

Within the last decade huge progress has been made regarding the development of new sources for the sustainable production of biomass, bioproducts and biofuels among which microalgae seem particularly promising [1,2]. Microalgae have a significant potential when compared to terrestrial crops, due to their higher productivity per surface and avoidance of competition for arable land needed for food production [3]. However, actual biomass productivity records are still far behind theoretical values [4]. Many projects therefore focus on how to improve growth yields and produce high value compounds [5,6]. As a model microalga, *Chlamydomonas reinhardtii* has been studied for over 50 years leading to the sequencing of its genome, comprehension of many of its vital cellular processes and significant progress in the setting up and improvement of genetic tools and techniques [7–12]. The green microalga is able to grow photoautrophically by using atmospheric CO_2, heterotrophically by using acetate and mixotrophically by using both carbon sources. The addition of acetate in the medium seems to be a good strategy to improve the production of high value compounds in the light [5]. *Chlamydomonas* is indeed able to metabolize this C_2 carbon through the glyoxylate cycle, a shunt of the tricarboxylic acid (TCA) cycle that was first discovered in 1958 [13] but only characterized recently in *Chlamydomonas* [14–16].

To gain further insight into the metabolic network allowing the improvement of growth by acetate, we took advantage of a null-mutant for isocitrate lyase (ICL), one of the two specific enzymes of the glyoxylate cycle [15]. This mutant is unable to metabolize acetate by the glyoxylate cycle and thus cannot grow in the dark [15]. We subjected this mutant to day/night cycles in a medium supplemented with acetate comparatively to its complemented strain containing a wild type copy of the isocitrate lyase gene and presenting full restoration of the phenotype [15,17]. In order to find genes/pathways whose importance varies between the two strains and which could thus be acetate-dependent, a transcriptomic analysis was performed. We chose to submit the cells to day/night cycles because these conditions mimic the open-air cultivation method. Such illumination conditions have recently gained more attention but only with minimal medium cultivation [18], which represents ideal conditions to reveal genes involved in cell cycle and diurnal rhythms but not those linked to organic carbon assimilation. RNA-seq data were collected every 4 hours during an entire day/night cycle (12h light/12h dark) and analysed by surprisal analysis. Surprisal analysis is a thermodynamic approach [19-20] which provides a biophysicochemical understanding and quantitative characterization of -omic expression levels using a molecule-centered approach [21]. Surprisal analysis defines a balance state, also called steady state, common to all the types of samples and additional constraints needed to quantify the deviation due to specific growth

conditions with respect to the balance state. It has been applied successfully for transcriptomic expression levels in human cell lines [21–24] and recently for metabolic data in *C. reinhardtii* [17].

Materials and Methods

1. Algal strains, cultivation conditions and sampling

ICL null-mutant strain (*icl*) and a complemented strain (*iclC*), both isolated and characterized in our laboratory [15], were cultivated in Tris-Acetate-Phosphate (TAP) medium. Cells were grown at 23°C in 12h light/12h dark with 100 $\mu E.m^{-2}.s^{-1}$ of cool white light during the day phase. Cells of both types (*icl* and *iclC*) were first adapted for three cycles in a multi-cultivator system (MC 1000-OD, Photon Systems Instruments) at initial concentration of about 0.5-1x10^6 cells. 9x10^7cells were harvested at the beginning of the third light phase when they were still in the exponential phase and re-suspended in 60 mL of fresh TAP medium (initial concentration of 1.5x10^6 cells/mL for both strains) in an identical multi-cultivator system. The relatively low initial concentration of 1.5x10^6 cells/mL allows a good distribution of the light inside the light-green cultures and an exponential growth phase for at least two consecutive 12h day/12h night cycles. Sampling was done over 28 hours at 4h intervals. The number of cells was counted using a Coulter counter (http://www.beckmancoulter.com). Samples were directly spun down by centrifugation and pellets were then frozen in liquid nitrogen and stored at -80°C until extraction. Experiments were carried out with three biological replicates for each strain. Acetate concentration in the medium was measured using a kit 'XXX'.

2. RNA extraction and sequencing

RNA from -80°C samples was isolated using Qiagen RNeasy Plant kit with a protocol adapted for *Chlamydomonas*. Library preparation started with 4µg total RNA for both complemented and mutant strains samples. Illumina Sequencing (PE 1x75 on a NextSeq500 machine) was performed at the GIGA-R Sequencing platform (University of Liège) following manufacturer's protocol (Illumina Inc, San Diego CA, USA).

3. Read trimming and quality filtering

Raw reads were analysed using FastQC (v.0.11.5) (www.bioinformatics.babraham.ac.uk/projects/) to evaluate read sequence quality.

Trimming and quality filtering of RNA-seq samples was done on the single-end reads using trimmomatic (v0.36) [25], removing low quality sequences (average Q20 over a 4-base sliding

window, minimum length of 50 bp with a leading and trailing quality threshold of Q25) **(Appendix 1)**. The trimmed reads were aligned to the *Chlamydomonas reinhardtii* genome v5.5 assembly [7] using STAR [26] with default presets except for intron size (-alignIntronMin 20 and -alignIntronMax 3000). The uniquely mapping reads were assigned to the transcripts using cuffquant and cuffdiff (v2.2.1) using a default fragment size of 200 (± 80) [27]. Normalization to library size and gene length was achieved by cufflinks and FPKM (Fragment Per Kilobase of transcript per Million mapped reads) values were obtained. PCA (Principal Component Analysis) on the expression values of the different replicates did not show any outliers among the replicates. FPKM values per samples averaged over the three independent biological replicates were computed using cuffdiff.

4. Surprisal analysis

All transcripts with an average FPKM value lower than 1 were removed from the data set, because most of the noise is due to low expression values [28]. In total 10873 genes were kept in the data set. Values lower than 0.01 FPKM were substituted with 0.01 FPKM to allow the computation of logarithms and expression ratios.

The natural logarithm (Y_{is}) of the K= 10873 gene expression values in each sample was subjected to surprisal analysis [21–24,29]. The constraints, $G_{i\alpha}$, and Lagrange multipliers, λ_α^s are determined via singular value decomposition (SVD) of the rectangular **Y** of the data as described in ref. [21].

$$\ln X_{is} = Y_{is} = \ln X_{is}^0 + \sum_{\alpha=1}^{N} G_{i\alpha}\lambda_\alpha^s \qquad (1)$$

Here α is the index of constraints, N is the total number of samples, i is the index of the gene and s is the index of the sample. The values for $G_{i\alpha}$ and λ_α^s are given by the SVD eigenvectors and the eigenvalues of the **Y** matrix:

$$G_{i\alpha} = U_{i\alpha} \text{ and } \lambda_\alpha^s = \omega_\alpha \left(\mathbf{V}^{\mathrm{T}}\right)_{\alpha s} \qquad (2)$$

where **U** and $\mathbf{V}^{\mathbf{T}}$ are respectively the left and right eigenvectors of the singular value decomposition and ω_α the singular values. The eigenvalues of the **Y** matrix are ordered by decreasing values and when all the N terms are kept, Eq. (1) is exact. Usually a few terms (smaller than the number of samples N) suffice to describe the input. Each constraint α corresponds to a given phenotype. For a given value of α, the surprisal analysis allows for a factorization between the weight of the constraint, $G_{i\alpha}$, on a given gene i and the Lagrange multiplier, λ_α^s, that is the weight of sample s in the phenotype that corresponds to the constraint α.

The first term in Eq. (1), $\ln X_{is}^{0} = G_{i0}\lambda_{0}^{s}$, corresponds to the balanced state. It is this stable state that is common to all the samples and with respect to which the changes in the gene expression levels due to the successive constraints, α=1, …, *N*, are expressed. By plotting the values of the Lagrange multipliers for the different samples for a given constraint α, one can identify different groups of samples that differ by the sign of their Lagrange multiplier, λ_{α}^{s}, in the phenotype α. The analysis of the weights of the corresponding vector, $G_{i\alpha}$, over the genes in terms of pathways gives access to the different pathway contributions to the phenotype under consideration.

5. Differential gene expression in the constraint vector $G_{i\alpha}$

Genes of the phenotype associated with each constraint α were ranked according the value of the weight $G_{i\alpha}$. According to this ranking, 100 smallest and largest values were considered differentially expressed for each phenotype. In the case of the balanced state, genes that correspond to a term, $G_{i0}\lambda_{0} > 0$ are the most stable and those for which $G_{i0}\lambda_{0} < 0$ are unstable. The latter are the genes that will appear with the largest and the smaller weights in the vectors $G_{i\alpha}$ of the constraints and therefore will be the most differentially expressed in the constraints, α=1, …, *N*.

6. Gene set enrichment

Genes were categorized in gene sets with KEGG pathways (http://www.genome.jp/kegg/pathway.html) using the functional annotation info for *C. reinhardtii* v5.5 (https://phytozome.jgi.doe.gov/pz/portal.html). K numbers were mapped to pathways included in the *C. reinhardtii* pathway maps (http://www.genome.jp/kegg-bin/show_organism?menu_type=pathway_maps&org=cre) using the KEGGREST package [30].

Pathways that correspond to a gene set with less than 10 genes were omitted from the dataset. For a given set of genes that corresponds to a pathway *J*, the $G_{i\alpha}$ values for genes that are respectively larger or smaller than zero were summed together to get respectively the positive (P) and negative (N) weight of the pathway *J* for constraint α:

$$P_{\alpha}^{J} = \sum_{i=1}^{S} G_{i\alpha}^{2} \text{ for } G_{i\alpha} > 0 \qquad (3)$$

$$N_{\alpha}^{J} = \sum_{i=1}^{S} G_{i\alpha}^{2} \text{ for } G_{i\alpha} < 0 \qquad (4)$$

The ratio

$$SR_{\alpha}^{J} = P_{\alpha}^{J} \Big/ N_{\alpha}^{J} \qquad (5)$$

is a measure for the contribution of the gene set of pathway *J* to constraint α. In Eqs. (3) and (4), *S* is the number of genes in pathway *J*.

Set ratios, SR_{α}^{J} were ordered according their value describing their importance for the described phenotype. These gene sets where all values $G_{i\alpha}$ are either positive or negative, were subsequently ranked on respectively P_{α}^{J} or N_{α}^{J}.

Both low ratios and high ratios are predicted by surprisal analysis to be important for the phenotype and to be enriched in their respective phenotypes.

Results

Two strains of *C. reinhardtii* were used in this experiment: a null-mutant deficient for isocitrate lyase and a complemented strain containing a wild-type copy of the gene, named *icl* and *iclC*, respectively. After three day/night cycles for adaptation, both strains were cultivated during 2 day/night cycles and sampling was realized during the first 28 hours at 8 time points (*t*=0, 4, 8, 12, 16, 20, 24 and 28). Since it is a 12h light/12h dark cycle, time points 0 and 24 (last point of the 12-hours darkness) and 4 and 28 (after 4 hours of illumination) are supposed to exhibit similar to identical values and trends.

1. Characteristics of the *icl* and *iclC* strains during the two day/night cycles

To evaluate how mutant *icl* cells response compared to complemented *iclC* cells in our cultivation conditions, we calculated the ratios between the cell numbers at the end and at the beginning of each illumination phase (**Table 1**). Whereas the *iclC* cells multiply more than 2 times during each light phase, the number of *icl* cells only increases by about 1.3-1.5, meaning that not all of them divide even once during the light phase. In addition, the cell ratio is about 1 in both dark phases for both strains. This means that *icl* cells are barely non-dividing during our experiment and that *iclC* cells are synchronized and divide during the light phase. We also provide in **Appendix 1** the growth curves of botho strains.

	Cycle 1				Cycle 2			
	Light phase (0-12h)		Dark phase (12-24h)		Light phase (24-36h)		Dark phase (36-48h)	
strain	*iclC*	*icl*	*iclC*	*icl*	*iclC*	*icl*	*iclC*	*icl*
cell ratio (±SD)	2.6 (±0.2)	1.5 (±0.0)	1.1 (±0.0)	1.0 (±0.0)	2.3 (±0.2)	1.3 (±0.0)	1.0 (±0.1)	0.9 (±0.1)

Table 1. Cells ratios (±SD) calculated between the end and the beginning of each illumination phase during two consecutive cycles. Mean calculated for 3 biological replicates per strain.

We next measured the acetate concentration of the medium at time point 24h and 48h in both strains. As shown in **Figure 1,** after the two day/night cycles, only 11.8% of the initial concentration is left in the medium in the control *icl*C strain while 85% is still present in the medium for the mutant *icl* strain. These results reflect the genetic differences between the two strains, since the control *iclC* strain is able to assimilate acetate by the glyoxylate cycle. Therefore our comparative transcriptomics analysis will allow deciphering the impact of the lack of the glyoxylate cycle when acetate is present in the medium. These results also show that some acetate is entering the mutant *icl* cells, the transcriptomics analyses will thus also allow detecting possible negative impacts of the acetate.

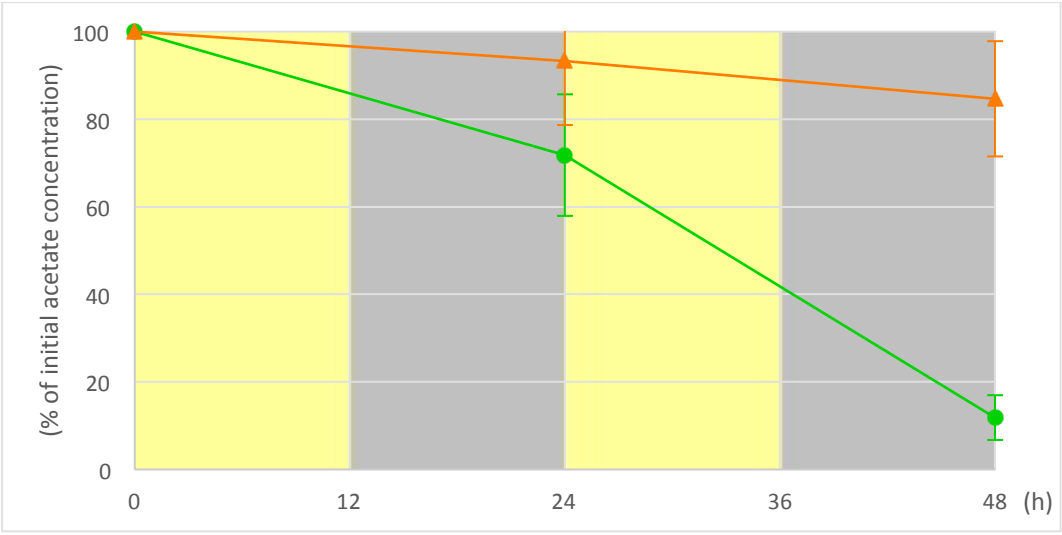


Figure 1. Evolution of the acetic acid concentration in the medium of *iclC* (green circles) and *icl* (orange triangles) strains (mean + SD of 3 biological replicates). Initial concentration (1g/L) is normalized to 100%. Yellow and grey backgrounds represent day and night phases, respectively.

The design of our experiment is relatively complex since data include and combine three variables: 8 consecutive time points, 2 different strains and 2 different light conditions (light vs dark). Since we aim to find constraints that would present a biological meaning for differentiation of samples over time, strain or light condition, we decided to use several approaches for data treatment with the surprisal methodology, with or without preliminary grouping of the data with respect to a specific variable.

2. Surprisal analysis applied to all the samples

We first performed the surprisal analysis on all the 16 samples (strain and time) together, meaning that the matrix Y used for the surprisal analysis is a 10883 x 16 matrix. The values of the Lagrange multipliers, $\lambda_{\alpha}^{s=icl,iclC}(t), \alpha = 0,1,2$ are plotted in **Figure 2**.

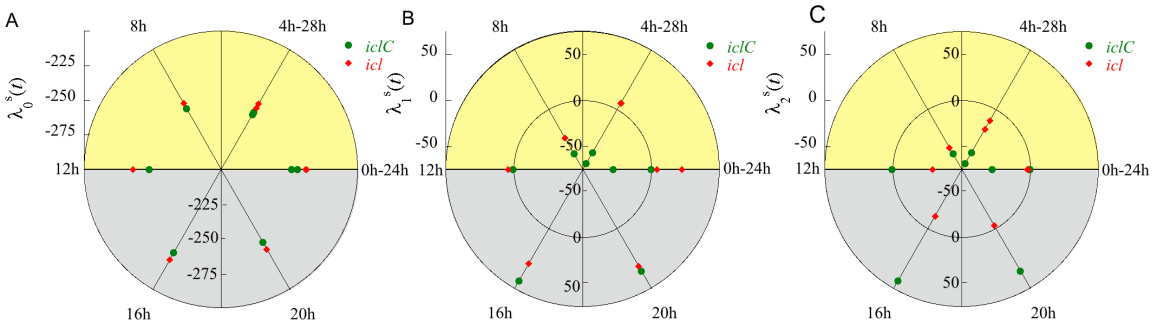


Figure 2. Values of the Lagrange multipliers $\lambda_{\alpha}^{s=icl,iclC}(t)$ as function of time for the complemented *s=iclC* (green) and mutant, *s=icl* (red) strains for an analysis of all the 16 samples in a single group. A. Balance state $\lambda_{0}^{s=icl,iclC}(t)$ B. First constraint $\lambda_{1}^{s=icl,iclC}(t)$ C. Second constraint $\lambda_{2}^{s=icl,iclC}(t)$. The day phase is indicated in yellow and the dark phase in grey. In panel B and C, negative values of the Lagrange multipliers fall inside the circle of 0 value. Time points 4h, 8h, 12h are in the day phase with 12h being the last point of the day, time points 16h, 20h 24h are in the dark phase, with time point 24h being the last one of the night. Time point 0h is the last point of the dark phase of the previous cycle; time point 28h is the day point of the next cycle.

Surprisal analysis determines Lagrange multipliers values $\lambda_{\alpha}^{icl,iclC}(t)$ which correspond to the weight of constraint α at time point *t* for the strain *icl* and for the strain *iclC* (see section "Surprisal analysis" in Materials and Methods). The constraints $\lambda_{\alpha}^{s}(t)$, α = 1, …,*N* prevent the intracellular entropy from reaching its global maximum. λ_0 corresponds to the balance state, or steady state, of the cells. The balance state is the reference stable distribution of expression levels common to all samples in the absence of any biological constraint. λ_1, λ_2 and further constraints have decreasing magnitudes, the higher the value of a Lagrange multiplier, the higher the weight of the corresponding phenotype.

The values of the Lagrange multipliers, $\lambda_0^{s=icl,iclC}(t)$, for the balance state (**Figure 2A**) are negative and much larger in absolute value than the values of the Lagrange multipliers of the other constraints (**Figure 2B, C**). They vary within a range of 30 units throughout the whole time range for both strains, which is expected for a steady state. When applying surprisal analysis to the data, we do not impose to λ_0 to be common to all the samples. The fact all samples have very similar values of $\lambda_0^s(t)$ confirms the existence of the balance state. The cyclic evolution of the values and the fact that data of time points 0 and 24 on one hand and 4 and 28 on the other are very close both suggest that the circadian regulation of the cell cycle can be observed.

Surprisal analysis also determines a gene transcript expression profile (G_α) associated with each constraint α. The components $G_{i\alpha}$ of the vector G_α determine the weight of transcript *i* in the phenotype associated with the constraint α. The Lagrange multiplier $\lambda_\alpha^s(t)$ determines the weight of the sample *s* in the phenotype α and can be different from sample to sample. One can rank the contribution of a transcript to a given phenotype according to its weight, $G_{i\alpha}$. As described in section 'Gene set enrichment' of Materials and Methods, the annotated genes [7] of *Chlamydomonas* are categorized in gene sets (KEGG pathways: Kyoto Encyclopedia of Genes and Genomes, http://www.kegg.jp/kegg/). This categorization therefore allows the identification of pathways that contribute most to the phenotype associated with a given constraint α via set ratio values SR_α^J (Eq. (5)) (see section "Gene set enrichment" in Materials and Methods) reflecting the contribution of a KEGG gene set *J* to the phenotype α for a given sample *s*. When for a given sample a KEGG pathway has a high set ratio, SR_α^J (Eq. (5) (P_α^J (positive sum) >> N_α^J (negative sum), Eqs. (3)-(4)), it means that it has a large weight in the constraint when Lagrange multipliers values, $\lambda_\alpha^s(t)$, are positive; when its set ratio is close to 0 (positive sum << negative sum), it has a large weight in the constraint when the corresponding Lagrange multipliers values are negative. The same reasoning is applied with large positive $G_{i\alpha}$ values of individual genes that have a high weight when the Lagrange multipliers are positive and large negative $G_{i\alpha}$ values having a high weight when the corresponding Lagrange multipliers are negative.

Out of the 10832 genes used in the surprisal analysis, 2710 could be categorized into 82 KEGG pathways (**Appendix 2**). In the top 10 pathways for the balance state (**Table 2**), essential physiological functions such as ribosomal activity, photosynthesis, mitochondrial respiration, carbon fixation and other central carbon metabolic pathways are found. These functions thus represent a group of cellular processes that are common to both strains in their steady state and whatever the

illumination condition. If the presence of “Glyoxylate and dicarboxylate metabolism” might be surprising at first view, since the glyoxylate cycle is absent in the mutant strain, it can be explained as follows: i) the glyoxylate metabolism is much wider than just the glyoxylate cycle. Glyoxylate is indeed also a key intermediate in other pathways such as photorespiration or purine degradation; ii) this KEGG pathway comprises 37 genes (34 were found in our genomic study) among which only a minority of 5 contributes to the glyoxylate cycle.

284

Balance state (α=0) – all 16 samples			
KEGG pathways	**Set size**	**Negative G subsetsize**	N_α^J
Ribosome	123	123	8.63E-04
Photosynthesis - antenna proteins	23	23	5.96E-04
Oxidative phosphorylation	66	66	3.85E-04
Carbon fixation in photosynthetic organisms	36	36	3.31E-04
Phagosome	29	29	3.21E-04
Glyoxylate and dicarboxylate metabolism	34	34	2.91E-04
Valine, leucine and isoleucine biosynthesis	12	12	2.74E-04
Citrate cycle (TCA cycle)	32	32	2.70E-04
2-Oxocarboxylic acid metabolism	30	30	2.65E-04
Lysine biosynthesis	10	10	2.36E-04

Table 2. KEGG pathways contributing most to the balance state of all the 16 samples in a single group.

Since λ_0 is negative (Figure 2A), the values of G_0 for the genes belonging to the most stable pathways are negative, leading to negative weights N_α^J (Eq. (4)).

In addition to ranking KEGG pathways according to their set ratio SR_α^J (Eq. (5)) values, $G_{i\alpha}$ values allow quantifying which individual genes contribute most to a given phenotype. Examination of the 100 genes that are the most highly expressed in the balance state confirms the results obtained by analysing the pathways since we can find in this list 67 ribosomal protein-coding transcripts (**Appendix 3, in bold**), 14 transcripts of proteins involved in the photosynthetic apparatus (PSA's and PSBR, LHCA8 and LHCBM's, PET's,…) (**Appendix 3, underlined**) and other ones linked with carbon metabolism such as GAP3.

The values of the Lagrange multipliers of the first two constraints exhibit a much larger range of variation than those of the balance state with 130 and 115 units for $\lambda_1^{icl,iclC}(t)$ and $\lambda_2^{icl,iclC}(t)$ respectively (**Figure 2B, C**). The values of the two Lagrange multipliers are of the same order of magnitude. Unfortunately, they do not allow us to strictly distinguish between two different phenotypes such as light versus dark phase or complemented versus mutant. This would have been the case if we had only positive values for one subset of the samples (samples of the light phase for example) and only negative values for the other subset (samples of the dark phase). Multipliers of the first constraint are of opposite sign when

comparing day and night time points for the complemented strain ($\lambda_1^{iclC}(t)$) (**Figure 2B**), but this is not the case for the values $\lambda_1^{icl}(t)$ of the mutant strain (**Figure 2B**) whose values are lower, preventing us to find a clear phenotype separation for the first constraint. The values of the Lagrange multipliers of the second constraint exhibits a rather similar (non-)trend behaviour, with maximum and minimum values peaking at the night-to-day and day-to-night transitions, respectively, but without any clear distinction of the sign of the Lagrange multipliers giving rise to a clear phenotype in terms of light and dark phases, or of the strain.

The KEGG pathways having the most positive or negative weight in the two constraints (**Tables 3 and 4**) are similar: 11 out of 20 (10 most positive and 10 most negative) of the first phenotype can also be found in the list for the second phenotype. This confirms that there is no specific interpretation of these two constraints in terms of the strain or the illumination variables. However, note that almost none of the top 10 KEGG pathways in the balance state (**Table 2**) can be found in the top 20 pathways (10 most positive and 10 most negative) of the phenotypes 1 and 2 (**Tables 3 and 4**), meaning that the two constraints do account for the differences in the measured gene expression levels with respect to the balance state. When experimental transcriptomic data are plotted (scatterplot) against computed data calculated with constraints 0, 1 and 2, we obtain a linear regression curve with a R^2 coefficient of 0.9251 (**Figure 3**), suggesting these constraints are sufficient to capture most of the experimental data. Further constraints will not be analysed since their amplitudes keep on decreasing and they have therefore less to none mathematical or biological meaning.

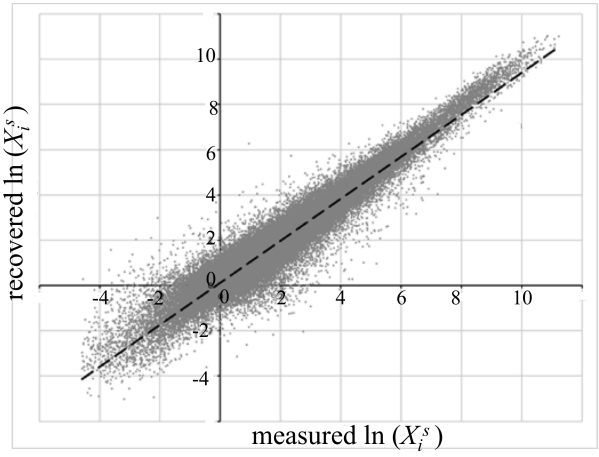


Figure 3. Scatter plot of the computed gene expression levels using the three main constraints α = 0, 1 and 2 vs. the measured experimental values. Linear regression curve is shown in dotted lines.

Constraint 1- Ten most positive pathways – all 16 samples together						
KEGG pathways	**Gene set size**	**Positive G subset size**	P_α^J	**Negative G subset size**	N_α^J	SR_α^J
DNA replication	27	27	5.94E-04	0	0	Inf
Mismatch repair	17	16	3.29E-04	1	5.21E-09	6.32E+04
RNA polymerase	26	24	1.11E-04	2	1.20E-08	9.31E+03
Ribosome biogenesis in eukaryotes	57	56	1.08E-04	1	1.16E-08	9.31E+03
Proteasome	34	31	5.64E-06	3	7.84E-09	7.20E+02
Base excision repair	16	14	2.87E-04	2	4.90E-07	5.87E+02
Homologous recombination	15	13	2.97E-04	2	8.26E-07	3.60E+02
Nucleotide excision repair	31	24	1,91E-04	7	1.78E-06	1.07E+02
Ubiquitin mediated proteolysis	55	44	6.26E-05	11	9.43E-07	6.64E+01
Arginine and proline metabolism	32	25	3.27E-05	7	1.02E-06	3.21E+01
Constraint 1 – Ten most negative pathways – all 16 samples together						
Photosynthesis antenna proteins	23	0	0	23	6.87E-04	0.00E+00
Carotenoid biosynthesis	13	1	6.79E-08	12	6.52E-05	1.04E-03
Porphyrin and chlorophyll metabolism	39	6	6.13E-06	33	2.90E-04	2.12E-02
Lysine biosynthesis	10	2	5.31E-07	8	1.79E-05	2.96E-02
Propanoate metabolism	17	3	6.81E-07	14	2.00E-05	3.40E-02
Terpenoid backbone biosynthesis	22	3	3.37E-06	19	9.30E-05	3.62E-02
Aminoacyl-tRNA biosynthesis	41	10	2.90E-06	31	3.95E-05	7.33E-02
Phenylalanine, tyrosine and tryptophan biosynthesis	21	3	2.72E-06	18	3.38E-05	8.04E-02
Fatty acid biosynthesis	20	3	4.07E-06	17	4.30E-05	9.46E-02
Biosynthesis of unsaturated fatty acids	16	7	5.07E-06	9	5.15E-05	9.84E-02

Table 3. KEGG pathways contributing most to the first constraint of all the 16 samples in a single group. See Eqs. (3-5) for the definition of P_α^J , N_α^J and SR_α^J respectively.

Constraint 2 – Ten most positive pathways – all 16 samples together						
KEGG pathways	**Gene set size**	**Positive G subset size**	P_α^J	**Negative G subset size**	N_α^J	SR_α^J
RNA polymerase	26	25	1.40E-04	1	2.25E-08	6.24E+03
Aminoacyl-tRNA biosynthesis	41	39	4.70E-05	2	1.45E-07	3.25E+02
Ribosome biogenesis in eukaryotes	57	53	1.85E-04	4	1.61E-06	1.15E+02
Spliceosome	107	93	3.36E-05	14	1.68E-06	2.00E+01
Phenylalanine, tyrosine and tryptophan biosynthesis	21	18	6.10E-05	3	3.14E-06	1.94E+01
Lysine biosynthesis	10	6	1.13E-04	4	6.21E-06	1.82E+01
Ribosome	123	114	6.52E-05	9	3.99E-06	1.64E+01
RNA degradation	54	42	5.42E-05	12	6.57E-06	8.26E+00
Carotenoid biosynthesis	13	11	8.11E-05	2	1.09E-05	7.42E+00
Steroid biosynthesis	11	7	1.49E-05	4	2.45E-06	6.09E+00
Constraint 2 – Ten most negative pathways – all 16 samples together						
Galactose metabolism	12	0	0	12	5.58E-05	0.00E+00
Mismatch repair	17	2	6.51E-08	15	5.87E-05	1.11E-03
DNA replication	27	5	1.31E-07	22	5.84E-05	2.25E-03
Regulation of autophagy	11	2	6.02E-07	9	1.01E-04	5.98E-03
Sphingolipid metabolism	16	3	2.68E-07	13	4.00E-05	6.69E-03
Photosynthesis - antenna proteins	23	7	2.58E-06	16	3.38E-04	7.63E-03
Valine, leucine and isoleucine degradation	20	3	2.30E-06	17	2.88E-04	8.01E-03
Arachidonic acid metabolism	12	2	3.98E-06	10	3.23E-04	1.23E-02
Proteasome	34	3	3.71E-07	31	2.96E-05	1.25E-02
Base excision repair	16	4	6.96E-07	12	3.63E-05	1.92E-02

Table 4. KEGG pathways contributing most to the second constraint of all the 16 samples in a single group. See Eqs. (3-5) for the definition of P_α^J , N_α^J and SR_α^J respectively.

3. Separate surprisal analysis for groups of samples

The transcriptomic data arise from the combination of several variables (two strains, light and dark conditions sampled by 8 time points) of the same importance. When surprisal analysis is carried out on all the 16 samples, the two largest Lagrange multipliers are of the same magnitude and a clear relation between the two constraints and a specific phenotype related to the nature of the strain or to the illumination condition does not emerge. In order to get more insight into the role of strain and of the illumination conditions, we therefore grouped the samples in two different ways before carrying out the surprisal analysis

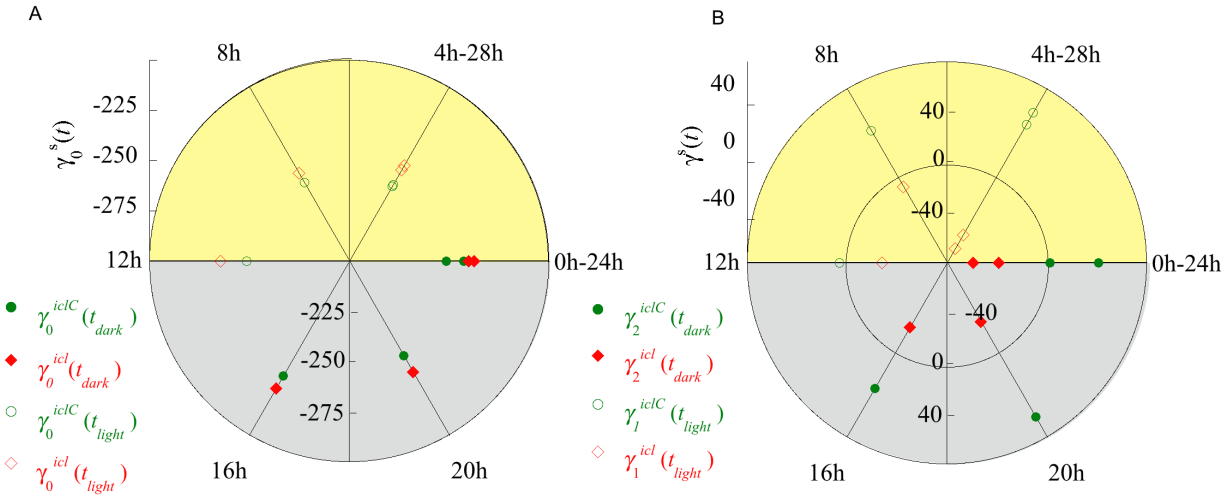
(i) according to the strain (mutant, *icl* or complemented, *iclC*), keeping the sample index to be light or dark phase.

(ii) according to the light or day phase, keeping the strain (*icl* or *iclC*) as a sample index

The surprisal analysis on the *icl* or the *iclC* samples (case (i)) separately yields results very similar to those obtained above, when all the samples are analysed together, and does not lead to light or dark phenotypes specific of the strain (data not shown). On the other hand, when the samples related to the day and the night phases are grouped together irrespective of the strain (case (ii)), one can clearly characterize specific phenotypes for the two strains. We discuss the results of the surprisal analysis for this grouping of the data in section 3 below.

4. Separate surprisal analysis for day and night phases

Surprisal analysis was applied to groups of 8 samples that include the two strains, but for the time points corresponding to the light phase (t= 4h, 8h, 12h and 28h) or for the time points corresponding to the dark phase (t = 0h, 16h, 20h and 24h) separately. In this case, we call the Lagrange multipliers $\gamma_{\alpha}^{s=icl,iclC}\left(t_{light}\right)$ and $\gamma_{\alpha}^{s=icl,iclC}\left(t_{dark}\right)$ to distinguish them from the case when the surprisal analysis is carried out on the 16 samples. The balance state and first or second constraint Lagrange multipliers from the "light" and "dark" analyses are plotted in **Figure 4A** and **B**, respectively.



394 **Figure 4.** Values of the Lagrange multipliers $\gamma_\alpha^{s=icl,iclC}\left(t_{light}\right)$ and $\gamma_\alpha^{s=icl,iclC}\left(t_{dark}\right)$ as function of time for
395 the complemented *s=iclC* (green) and mutant, *s=icl* (red) strains when light and dark samples are
396 analysed separately. The Lagrange multipliers for the group of the dark phase samples appear with filled
397 symbols and those for the group of the light phase samples appear with empty symbols. A. The Lagrange
398 multipliers of the balance state for the analysis of the dark phase $\gamma_0^{s=icl,iclC}\left(t_{dark}\right)$ and for the analysis of
399 the light phase, $\gamma_0^{s=icl,iclC}\left(t_{light}\right)$ B. First $\gamma_1^{s=icl,iclC}\left(t_{light}\right)$ and second $\gamma_2^{s=icl,iclC}\left(t_{dark}\right)$ constraints. The period
400 corresponding to the day phase is indicated in yellow and the one corresponding to the dark phase in
401 grey.

402

403 As in the previous cases, the γ_0 values of the steady state barely vary within a 25-30 units range in both
404 analyses (**Figure 4A**) and are similar for the dark phase group and for the light phase group. 8 out of 10
405 top pathways are common to the balance states of the dark and light groups of samples (**Table 5**) and
406 also common with the balance state of the "16 samples together" analysis. However, some stable
407 pathways are different in the balance state of light and dark samples. The "Photosynthesis – antenna
408 proteins" pathway is ranked in 2nd position in the balance state of the light samples while it logically only
409 appears in 46th position in the balance state of the dark ones. This reveals that the function associated
410 with the "Photosynthesis – antenna proteins" pathway is important in the balance phase of the light
411 samples and is not prevailing in the balance state of the dark samples. Conversely "Citrate cycle"
412 pathways appears on 7th position in the analysis of the dark phase but only on 49th in the analysis of light

ones emphasizing the importance of this cycle during the night period when cells cannot rely anymore on photosynthesis to produce ATP molecules sustaining their growth.

Light samples – balance state α= 0			
KEGG pathways	**Gene set size**	**Negative G subsetsize**	N_α^-
Ribosome	123	123	8.04E-04
Photosynthesis - antenna proteins	23	23	7.65E-04
Oxidative phosphorylation	66	66	3.87E-04
Carbon fixation in photosynthetic organisms	36	36	3.43E-04
Phagosome	29	29	3.14E-04
Glyoxylate and dicarboxylate metabolism	34	34	2.92E-04
2-Oxocarboxylic acid metabolism	30	30	2.62E-04
Valine, leucine and isoleucine biosynthesis	12	12	2.59E-04
Pentose phosphate pathway	24	24	2.57E-04
Lysine biosynthesis	10	10	2.30E-04
Dark samples balance state α= 0			
Ribosome	123	123	9.01E-04
Oxidative phosphorylation	66	66	3.70E-04
Phagosome	29	29	3.17E-04
Carbon fixation in photosynthetic organisms	36	36	3.07E-04
Valine, leucine and isoleucine biosynthesis	12	12	2.80E-04
Glyoxylate and dicarboxylate metabolism	34	34	2.80E-04
Citrate cycle (TCA cycle)	32	32	2.60E-04
2-Oxocarboxylic acid metabolism	30	30	2.60E-04
Lysine biosynthesis	10	10	2.34E-04
Alanine, aspartate and glutamate metabolism	26	26	2.16E-04

Table 5. KEGG pathways contributing most to the balance state when light and dark samples are analysed separately. Since γ_0 is negative (Figure 3A), the values of G_0 for the genes belonging to the most stable pathways are negative, leading to negative weights N_α^J (Eq. (4)).

The next two constraints have basically identical orders of magnitude, as was the case of the analysis of the 16 samples together. However, here, the first constraint in the case of the light samples and the second one in the case of the dark ones allow characterizing the genomic response of the two strains to illumination conditions. Both in the light group of data as well as dark group we can notice a change of sign of the Lagrange multiplier between samples obtained with mutant and complemented strains, meaning that there is a phenotypic difference between the two strains in both illumination conditions. The analysis of the second constraint in the case of the light samples and of the first one in the case of the dark ones did not lead to any clear phenotype identification.

Analysis of constraint $\gamma_1^{s=icl,iclC}\left(t_{light}\right)$ *of light-phase samples*

In the analysis of the group of light-phase samples all Lagrange multipliers values $\gamma_1^{s=icl,iclC}\left(t_{light}\right)$ are positive for the complemented strain and can be linked with pathways related to carbon metabolism and essential compounds such as amino acids, fatty acids, pyruvate or propanoate (**Table 6**). The presence of "Peroxisome" in this top 10 is also of particular interest because it reflects the genotypic difference between the two strains since we have previously shown that peroxisomal microbodies are the main place for glyoxylate cycle in *C. reinhardtii* [16]: acetate assimilation has an important role and influence in complemented cells.

Light samples – constraint 1 – 10 most positive pathways- *iclC* phenotype				**Light samples – constraint 1 – 10 most negative pathways – *icl* phenotype**			
KEGG pathways	P_α^J	N_α^J	SR_α^J	**KEGG pathways**	P_α^J	N_α^J	SR_α^J
Fatty acid degradation	5.94E-05	1.14E-06	5.23E+01	DNA replication	5.80E-07	4.00E-04	1.45E-03
Propanoate metabolism	7.03E-05	2.27E-06	3.10E+01	Mismatch repair	8.70E-07	2.08E-04	4.18E-03
Lysine biosynthesis	2.77E-05	9.08E-07	3.05E+01	Regulation of autophagy	5.33E-07	3.61E-05	1.48E-02
Terpenoid backbone biosynthesis	1.64E-05	8.44E-07	1.95E+01	Base excision repair	2.36E-06	1.57E-04	1.50E-02
Pyruvate metabolism	9.16E-05	5.45E-06	1.68E+01	Homologous recombination	4.08E-06	2.43E-04	1.68E-02

Carotenoid biosynthesis	2.55E-05	2.01E-06	1.27E+01	SNARE interactions in vesicular transport	4.06E-07	1.24E-05	3.29E-02
Citrate cycle (TCA cycle)	8.46E-05	6.93E-06	1.22E+01	Endocytosis	8.93E-07	2.62E-05	3.40E-02
Phenylalanine, tyrosine and tryptophan biosynthesis	3.99E-05	3.42E-06	1.16E+01	Ubiquitin mediated proteolysis	2.43E-06	4.80E-05	5.06E-02
Peroxisome	4.34E-05	3.89E-06	1.12E+01	Nucleotide excision repair	7.03E-06	1.20E-04	5.87E-02
Biosynthesis of unsaturated fatty acids	2.71E-05	2.61E-06	1.04E+01	Pentose and glucuronate interconversions	2.17E-06	2.72E-05	7.96E-02

Table 6. KEGG pathways contributing most to the first constraint when light and dark samples are analysed separately (8 light samples considered). See Eqs. (3-5) for the definition of P_α^J , N_α^J and SR_α^J respectively.

The importance of acetate metabolization is also particularly noticed in the top 100 genes list (**Appendix 4**). Transcripts of 4 isoforms of the glyoxylate cycle enzymes (MAS1, 5th position; ICL1, 7th; ACS3, 12th and CIS2, 32nd) are found next to PCK1 (22nd), encoding the enzyme catalysing the first step of gluconeogenesis, and AST1 (25th) encoding the enzyme which catalyses the reversible interconversion of oxaloacetate and glutamate to aspartate and α-ketoglutarate. The transcripts of two enzymes of the TCA cycle, IDH3 (33rd) that catalyses the decarboxylation of isocitrate and, two steps further, succinate-CoA ligase (91st) that catalyses the conversion of succinyl-CoA to succinate are also found (**Figure 5**). We want to mention again at this point that 68 up to 74/100 genes in the lists of these third approach have unknown functions.

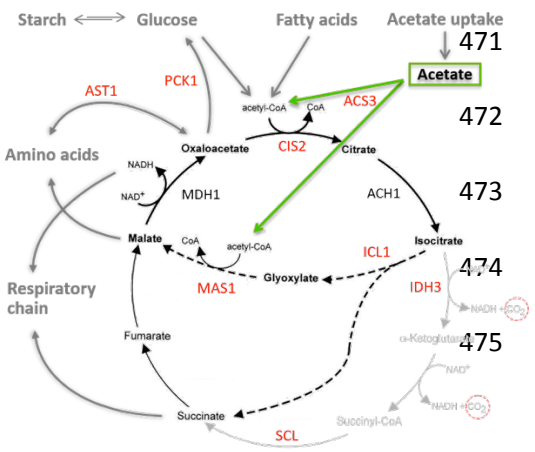


Figure 5. Glyoxylate cycle and its connexions with tricarboxylic acid cycle and central carbon metabolism in *C. reinhardtii*. The glyoxylate cycle is a shunt of the TCA cycle and avoids the two decarboxylation steps of the TCA (in grey). Enzyme isoforms in red are encoded by transcripts found in the analysis of constraint $\gamma_1(t)$ of light-phase samples for the complemented strain.

Regarding the mutant phenotype in the light, pathways related to cell cycle/division and stress conditions are highly ranked and thus important (**Table 6**). It is rather understandable to find functions such as "Regulation of autophagy" or "Ubiquitin mediated proteolysis" in this top 10 since mutant cells cannot develop correctly under these 12h light/12h dark conditions and are somehow stressed by the presence of acetate in the medium [15]. However, it is more intriguing to see that some cell division functions ("DNA replication", "Mismatch repair", "Base excision repair", "Homologous recombination", "Nucleotide excision repair") also contribute to the phenotype. This is in contrast with the complemented cells where these processes are found important in constraint 2 of the dark phase (see below). This suggests that mutant cells activate the cellular machinery to divide during the light phase and not during the dark phase (**Table 1**). We also found genes related to cell division in the list of the 100 genes contributing most to the *icl* phenotype in the light phase (**Appendix 4**).

Analysis of constraint $\gamma_2^{s=icl,iclC}\left(t_{dark}\right)$ *of dark-phase samples*

In the analysis of the samples of the dark phase, all Lagrange multipliers of constraint 2 are positive for the complemented strain (*iclC*) (**Figure 4B**). The constraint is characterized by a high positive weight of the cell cycle/division functions ("RNA polymerase", "DNA replication", "Homologous recombination",…) and protein processing ("Proteasome", "Protein export" and "Spliceosome") (**Table 7)**. Therefore cells of the *iclC* strain prepare the division during the dark phase and divide during the light phase (see **Table 1**). We have already observed that *iclC* cells divide synchronously at the beginning of the day phase in TAP medium provided that the temperature is maintained at 23°C (Willamme et al. 2015). Although the 12h light/12h dark cultivation conditions are not exactly the same as ours since the cells are maintained under turbidostatic control at 28°C and 125 $\mu E.m^{-2}.s^{-1}$ during the light phase, Zones et al. (2015) (18) observed that cells cultivated in a medium without acetate prepare the division during the light phase and divide in the dark phase, the hatching of the daughter cells occurring at the middle of the dark phase. This pattern is thus in contrast to what we observe with the *iclC* strain cultivated with acetate and somewhat similar to that of the *icl* strain. Indeed the surprisal analysis shows that cell division pathways are important during the light phase for the mutant strain. However, cells cannot divide during the dark phase, probably because of the lack of energy. Thus, for this particular physiological event, the *icl* mutant could behave as it were in a minimal medium.

Light samples – constraint 1 – 10 most positive pathways- *iclC* phenotype				Light samples – constraint 1 – 10 most negative pathways – *icl* phenotype			
KEGG pathways	P_α^J	N_α^J	SR_α^J	**KEGG pathways**	P_α^J	N_α^J	SR_α^J
Proteasome	6.64E-05	0.00E+00	inf	DNA replication	2.11E-08	2.02E-04	1.04E-04
RNA polymerase	6.05E-05	0.00E+00	inf	Mismatch repair	2.78E-06	1.51E-04	1.84E-02
DNA replication	9.00E-05	1.82E-07	4.95E+02	Regulation of autophagy	1.42E-06	7.14E-05	1.99E-02
Homologous recombination	3.98E-05	5.72E-07	6.94E+01	Base excision repair	6.45E-06	5.76E-05	1.12E-01
N-Glycan biosynthesis	8.32E-05	2.63E-06	3.16E+01	Homologous recombination	2.68E-05	1.86E-04	1.44E-01
Mismatch repair	4.77E-05	1.53E-06	3.11E+01	SNARE interactions in vesicular transport	4.05E-05	2.74E-04	1.48E-01
Protein export	5.47E-05	3.07E-06	1.78E+01	Endocytosis	7.66E-06	4.16E-05	1.84E-01
Nucleotide excision repair	3.23E-05	1.95E-06	1.65E+01	Ubiquitin mediated proteolysis	3.02E-05	1.61E-04	1.88E-01
Spliceosome	7.18E-05	4.92E-06	1.56E+01	Nucleotide excision repair	3.99E-06	1.75E-05	2.28E-01
mRNA surveillance pathway	4.66E-05	3.50E-06	1.33E+01	Pentose and glucuronate interconversions	1.19E-05	4.48E-05	2.66E-01

509

510 **Table 7**. KEGG pathways contributing most to the second constraint when light and dark samples are
511 analysed separately (8 dark samples considered). See Eqs. (3-5) for the definition of P_α^J , N_α^J and SR_α^J
512 respectively.

513

514 The presence of 10 transcripts (**Appendix 5, in bold**) linked with cell/organelle division in the top 100
515 genes also reflects this *iclC* dark phenotype. Like in the light phase, *ICL1*, the gene coding for isocitrate
516 lyase, is in 4th position of the list but is not accompanied by the transcripts encoding enzymes of the
517 glyoxylate cycle. We can also find in this list FDX5 (2nd), encoding an apoferredoxin that displays physical
518 interaction with fatty acid desaturases to maintain membrane structure and dark metabolism [32] and
519 TAL1 (65th) and GAP1 (91st) that encode enzymes that both have glyceraldehyde 3-phosphate as
520 substrate, a key intermediate between glycolysis, gluconeogenesis and pentose phosphate pathway.

521 Finally, the phenotype of mutant cells (*icl*) samples in the dark (**Table 7**) reveals that "Regulation of
522 autophagy" is at the very first position of the most negative pathways, which reflects that cells suffer

from their inability to assimilate acetate provided in the cultivation medium. This is also reflected in the genes list with the presence of 6 transcripts linked to limited carbon availability (**Appendix 6, in bold**), and 6 transcripts related to stressful conditions response (**Appendix 6, underlined**). If understanding the presence of the other pathways and genes is less straightforward at first sight we think they can actually be related. Mutant cells lacking the glyoxylate cycle function have to degrade their own carbon components such as starch, fatty acids and amino for acetyl-CoA (**Figure 5**) and energy release during the night. Consequently, we could find pathways (**Table 7**) and genes related to these functions (**Appendix 6**). Acetyl-CoA production is for example reflected through "beta-Alanine metabolism" pathway: indeed beta-alanine is a component of pantothenic acid (known as vitamin B_5) and thus of coenzyme A ("Pentothenate acid and CoA biosynthesis"), which itself is crucial for the synthesis of acetyl-CoA. Fatty acids metabolism is exemplified by the presence of "Arachidonic acid metabolism", a polyunsaturated fatty acid that can be found in the phospholipids of cellular membranes, "alpha-Linolenic acid metabolism", also a polyunsaturated fatty acid and "Propanoate metabolism". Moreover the transcript MCC1 encodes a methylcrotonyl-CoA carboxylase, (**Appendix 6**), a mitochondrial enzyme essential for the breakdown of amino acids such as leucine and isoleucine ("Valine, leucine and isoleucine degradation"). Altogether our results suggest a rallying of available carbon source in the cells towards catabolism. This global metabolic remodelling is also supported by the presence of the "ABC transporters" pathway and 6 transcripts encoding transporters (**Appendix 6, in italics**).

Conclusions

Surprisal analysis was applied for the first time to transcriptomic data of *Chlamydomonas reinhardtii* obtained from a complex experimental design with 2 different strains, 2 illumination conditions and 8 consecutive time points. The aim of this analysis was to assess whether the constraints could identify and reveal specific phenotypes under day/night cyclic cultivation conditions that are strain dependent. When surprisal analysis is applied to all the 16 samples, two Lagrange multipliers have a larger value but the corresponding phenotypes do not allow distinguishing a strain-dependent specific response to light or dark phases. When surprisal analysis is carried out separately on two groups of 8 samples of the same strain (for which the light and dark phases are the variable), it is also not possible to distinguish a clear illumination dependent phenotype for the two strains.

However, when surprisal analysis was applied on day and night groups of 8 samples that each gather the two different strains, we could clearly distinguish a phenotype linked to the strain dependent response to light and dark phases. Both in the light phase group and in the dark phase one, the Lagrange

multipliers of the *ic*l and *iclC* strains have opposite signs. Complemented, *iclC*, samples are highly influenced by acetate assimilation during the light phase and by cellular replication during the night phase. On the other hand, mutant, *icl*, cells are also influenced by similar cellular replication pathways but during the day period. Stress-related pathways and genes have important weights on the mutant phenotype during both phases and the night period is also marked by low-CO_2 response transcripts highlighting that these cultivation conditions and especially the night phase of the cycle is a very stressful period for this isocitrate lyase mutant that can only rely on photosynthesis to fix carbon in the light and depends on its own energy stocks during the night period.

Through the three different ways of grouping the data before carrying out surprisal analysis we could obtain each time a stable balance state that showed very similar pathways and genes whatever the grouping with only few variation (absence of "Photosynthesis – antenna proteins" in the top 10 most positive pathways of dark group analysis, for example). Therefore, despite the fact that the data were obtained from different strains, illumination conditions or time points, surprisal analysis allowed us to describe a steady state that is common to all the samples of the experiment.

Lastly, we think it is useful to notice the large number of genes with unknown functions present in the most influential genes lists (up to 80 out of 100). This emphasizes that many (functions of) genes that play important roles on cells growth and development and especially on circadian rhythms have still to be discovered and/or studied. Using these gene lists as well as KEGG pathways that comprise 2710 genes in our analysis, in comparison with the 17741 genes identified in the genome of *C. reinhardtii*, is of course very useful to assign biological meaning to mathematical data but this still has some limitations and should be further enhanced in order to make it an even more powerful tool. However, this also makes particularly relevant and significant the presence of genes with known functions in those lists.

Acknowledgements

RW is recipient of FRIA (Fonds pour la Recherche dans l'Industrie et l'Agriculture); FR is Research Director of FRS-FNRS (Fonds National de la Recherche Scientifique). All the authors gratefully acknowledge the support of FP7-Future and Emerging Technologies-Open Project- BAMBI 618024; CR acknowledges FRS-FNRS (CDR J.0265.17). The funding bodies did not have a role in the design of the study, data collection, analysis, interpretation of data, writing the manuscript, nor the decision to publish.

List of appendices

Appendix 1. Reads information for each sample.

Appendix 2. KEGG pathways defined for *C. reinhardtii* (http://www.kegg.jp/kegg/). Names of the pathways (ko indexes), number of genes contained in the pathway and number of genes of the pathway detected in our transcriptomic data are mentioned.

Appendix 3. List of the 100 most highly expressed genes in the balance constraint when all 16 samples are grouped together. *C. reinhardtii* genome version v5.5. #N/A means the function of the gene is unknown.

Appendix 4. List of the 100 most positive expressed genes (*iclC* phenotype) of the first constraint when 8 light samples are analysed separately. *C. reinhardtii* genome version v5.5. #N/A means the function of the gene is unknown.

Appendix 5. List of the 100 most positive expressed genes (*iclC* phenotype) of the second constraint when 8 dark samples are analysed separately. *C. reinhardtii* genome version v5.5. #N/A means the function of the gene is unknown.

Appendix 6. List of the 100 most negative expressed genes (*icl* phenotype) of the second constraint when dark samples are analysed separately. *C. reinhardtii* genome version v5.5. #N/A means the function of the gene is unknown.

Bibliography

[1] R.H. Wijffels, O. Kruse, K.J. Hellingwerf, Potential of industrial biotechnology with cyanobacteria and eukaryotic microalgae, Curr. Opin. Biotechnol. 24 (2013) 405–413. doi:10.1016/j.copbio.2013.04.004.

[2] V. Klassen, O. Blifernez-Klassen, L. Wobbe, A. Schlüter, O. Kruse, J.H. Mussgnug, Efficiency and biotechnological aspects of biogas production from microalgal substrates, J. Biotechnol. 234 (2016) 7–26. doi:10.1016/j.jbiotec.2016.07.015.

[3] R.H. Wijffels, M.J. Barbosa, An Outlook on Microalgal Biofuels, Science 329 (2010) 796–799. doi:10.1126/science.1189003.

[4] P.M. Schenk, S.R. Thomas-Hall, E. Stephens, U.C. Marx, J.H. Mussgnug, C. Posten, et al., Second Generation Biofuels: High-Efficiency Microalgae for Biodiesel Production, BioEnergy Res. 1 (2008) 20–43. doi:10.1007/s12155-008-9008-8.

[5] K.J. Lauersen, O. Kruse, J.H. Mussgnug, Targeted expression of nuclear transgenes in Chlamydomonas reinhardtii with a versatile, modular vector toolkit, Appl. Microbiol. Biotechnol. 99 (2015) 3491–3503. doi:10.1007/s00253-014-6354-7.

[6] G. Benvenuti, J. Ruiz, P.P. Lamers, R. Bosma, R.H. Wijffels, M.J. Barbosa, Towards microalgal triglycerides in the commodity markets, Biotechnol. Biofuels. 10 (2017) 188. doi:10.1186/s13068-

622 017-0873-2.

623 [7] S.S. Merchant, S.E. Prochnik, O. Vallon, E.H. Harris, S.J. Karpowicz, G.B. Witman, et al., The
624 Chlamydomonas Genome Reveals the Evolution of Key Animal and Plant Functions, Science 318
625 (2007) 245–252.

626 [8] S. Eberhard, G. Finazzi, F.-A. Wollman, The Dynamics of Photosynthesis, Annu. Rev. Genet. 42
627 (2008) 463–515. doi:10.1146/annurev.genet.42.110807.091452.

628 [9] T. Schulze, K. Prager, H. Dathe, J. Kelm, P. Kießling, M. Mittag, How the green alga
629 Chlamydomonas reinhardtii keeps time, Protoplasma. 244 (2010) 3–14. doi:10.1007/s00709-010-
630 0113-0.

631 [10] M. Siaut, S. Cuiné, C. Cagnon, B. Fessler, M. Nguyen, P. Carrier, et al., Oil accumulation in the
632 model green alga Chlamydomonas reinhardtii: characterization, variability between common
633 laboratory strains and relationship with starch reserves., BMC Biotechnol. 11 (2011) 7.
634 doi:10.1186/1472-6750-11-7.

635 [11] I.K. Blaby, C.E. Blaby-Haas, N. Tourasse, E.F.Y. Hom, D. Lopez, M. Aksoy, et al., The
636 Chlamydomonas genome project: A decade on, Trends Plant Sci. 19 (2014) 672–680.
637 doi:10.1016/j.tplants.2014.05.008.

638 [12] J.H. Mussgnug, Genetic tools and techniques for Chlamydomonas reinhardtii, (2015) 5407–5418.
639 doi:10.1007/s00253-015-6698-7.

640 [13] H.L. Kornberg, A.M. Gotto, P. Lund, Effect of growth substrates on isocitratase formation by
641 Pseudomonas ovalis Chester, Nature. 182 (1958) 1430–1431.
642 doi:http://dx.doi.org/10.1038/1821430a0.

643 [14] Y. Hayashi, A. Shinozaki, Visualization of microbodies in Chlamydomonas reinhardtii, J. Plant Res.
644 125 (2012) 579–586. doi:10.1007/s10265-011-0469-z.

645 [15] C. Plancke, H. Vigeolas, R. Höhner, S. Roberty, B. Emonds-Alt, V. Larosa, et al., Lack of isocitrate
646 lyase in Chlamydomonas leads to changes in carbon metabolism and in the response to oxidative
647 stress under mixotrophic growth, Plant J. 77 (2014) 404–417. doi:10.1111/tpj.12392.

648 [16] K.J. Lauersen, R. Willamme, N. Coosemans, M. Joris, O. Kruse, C. Remacle, Peroxisomal
649 microbodies are at the crossroads of acetate assimilation in the green microalga Chlamydomonas
650 reinhardtii, Algal Res. 16 (2016) 266–274. doi:10.1016/j.algal.2016.03.026.

651 [17] R. Willamme, Z. Alsafra, R. Arumugam, G. Eppe, F. Remacle, R.D. Levine, et al., Metabolomic
652 analysis of the green microalga Chlamydomonas reinhardtii cultivated under day/night

conditions, J. Biotechnol. 215 (2015) 20–26. doi:10.1016/j.jbiotec.2015.04.013.

[18] J.M. Zones, I.K. Blaby, S.S. Merchant, J.G. Umen, High-Resolution Profiling of a Synchronized Diurnal Transcriptome from Chlamydomonas reinhardtii Reveals Continuous Cell and Metabolic Differentiation, Plant Cell 27 (2015) 2743–2769. doi:10.1105/tpc.15.00498.

[19] R.D. Levine, Information theory approach to molecular reaction dynamics. Annu. Rev. Phys. Chem. 29 (1978) 59–92.

[20] R.D. Levine Molecular Reaction Dynamics (2005, Cambridge Univ Press, Cambridge, UK).

[21] F. Remacle, N. Kravchenko-Balasha, A. Levitzki, R.D. Levine, Information-theoretic analysis of phenotype changes in early stages of carcinogenesis., Proc. Natl. Acad. Sci. U. S. A. 107 (2010) 10324–9. doi:10.1073/pnas.1005283107.

[22] N. Kravchenko-Balasha, F. Remacle, A. Gross, V. Rotter, A. Levitzki, R.D. Levine, Convergence of Logic of Cellular Regulation in Different Premalignant Cells by an Information Theoretic Approach, BMC Syst. Biol. 5 (2011) 42. doi:10.1186/1752-0509-5-42.

[23] N. Kravchenko-Balasha, A. Levitzki, A. Goldstein, V. Rotter, A. Gross, F. Remacle , R.D. Levine, On a fundamental structure of gene networks in living cells, Proc. Natl. Acad. Sci. U.S.A. 109 (2012) 4702–4707. doi:10.1073/pnas.1200790109/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1200790109.

[24] S. Zadran, F. Remacle, R.D. Levine, miRNA and mRNA cancer signatures determined by analysis of expression levels in large cohorts of patients, Proc. Natl. Acad. Sci. U. S. A. 110 (2013) 19160–19165. doi:10.1073/pnas.1316991110.

[25] A.M. Bolger, M. Lohse, B. Usadel, Trimmomatic: A flexible trimmer for Illumina sequence data, Bioinformatics. 30 (2014) 2114–2120. doi:10.1093/bioinformatics/btu170.

[26] A. Dobin, C.A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, et al., STAR: Ultrafast universal RNA-seq aligner, Bioinformatics. 29 (2013) 15–21. doi:10.1093/bioinformatics/bts635.

[27] C. Trapnell, A. Roberts, L. Goff, G. Pertea, D. Kim, D.R. Kelley, et al., TopHat and Cufflinks_Protocol, Nat. Protoc. 7 (2012) 562–78. doi:10.1038/nprot.2012.016.

[28] V. Piras, K. Selvarajoo, The reduction of gene expression variability from single cells to populations follows simple statistical laws, Genomics. 105 (2015) 137–144. doi:10.1016/j.ygeno.2014.12.007.

[29] F. Remacle, R.D. Levine, Statistical thermodynamics of transcription profiles in normal development and tumorigeneses in cohorts of patients, Eur. Biophys. J. 44 (2015) 709–726.

[30] D. Tenenbaum, KEGGREST: Client-side REST access to KEGG. R package version 1.12.2., (2016). doi:10.18129/B9.bioc.KEGGREST.

[31] K. Matsumura, T. Yagi, A. Hattori, M. Soloviev, K. Yasuda, Using single cell cultivation system for on-chip monitoring of the interdivision timer in Chlamydomonas reinhardtii cell cycle., J. Nanobiotechnology 8 (2010) 23. doi:10.1186/1477-3155-8-23.

[32] W. Yang, T.M. Wittkopp, X. Li, J. Warakanont, A. Dubini, C. Catalanotti, et al., Critical role of Chlamydomonas reinhardtii ferredoxin-5 in maintaining membrane structure and dark metabolism, Proc. Natl. Acad. Sci. U. S. A. 112 (2015) 14978–83. doi:10.1073/pnas.1515240112.

	Identifier	Sample	Time	Run1	Run2	Run3	Total trimmed reads	Uniquely mapped reads	
E570	S1	icl-1	0	7,217,147	9,986,544	8,899,546	24,594,592	10,889,198	44.27%
E571	S2	icl-2	0	6,985,359	9,593,406	8,567,402	23,672,579	5,397,631	22.80%
E572	S3	icl-3	0	8,100,615	11,034,273	9,882,926	27,233,161	16,728,413	61.43%
E573	S4	icl-1	4	6,467,075	9,146,339	8,745,481	22,827,602	6,414,517	28.10%
E574	S5	icl-2	4	7,383,804	10,553,381	10,073,812	26,328,898	5,172,493	19.65%
E575	S6	icl-3	4	6,743,101	9,431,694	9,013,797	23,672,337	8,634,465	36.47%
E576	S7	icl-1	8	6,247,151	8,500,819	7,622,941	20,935,847	9,267,671	44.27%
E577	S8	icl-2	8	6,561,665	9,096,924	8,113,833	22,299,466	11,949,621	53.59%
E578	S9	icl-3	8	6,798,083	9,320,352	8,290,004	22,864,576	12,424,631	54.34%
F077	S10_merged	icl-1	12	32,755,414			29,897,157	28,340,363	94.79%
F078	S11_merged	icl-2	12	33,261,222			30,053,211	28,237,858	93.96%
F079	S12_merged	icl-3	12	34,400,218			31,270,607	29,713,890	95.02%
F080	S13_merged	icl-1	16	31,882,527			28,375,109	26,673,053	94.00%
F081	S14_merged	icl-2	16	34,901,956			31,731,913	26,074,281	82.17%
F082	S15_merged	icl-3	16	50,662,389			45,946,899	43,403,802	94.47%
E585	S16	icl-1	20	7,284,467	9,977,663	8,904,835	24,456,403	15,934,392	65.15%
E586	S17	icl-2	20	7,200,127	10,004,412	8,931,750	24,432,043	13,946,239	57.08%
E587	S18	icl-3	20	7,441,326	10,105,891	9,078,986	24,839,544	15,642,532	62.97%
E588	S19	icl-1	24	7,311,700	10,128,151	8,994,962	24,857,333	10,220,754	41.12%
E589	S20	icl-2	24	6,731,836	9,379,166	8.373.664	22,833,919	9,439,133	41.34%
E590	S21	icl-3	24	7,008,238	9,563,936	8,602,426	23,498,155	6,177,451	26.29%
E591	S22	icl-1	28	6,844,960	9,360,993	8,320,745	23,089,569	5,775,659	25.01%
E592	S23	icl-2	28	7,056,772	9,826,967	8,745,082	24,069,072	6,828,397	28.37%
E593	S24	icl-3	28	7,315,756	10,170,840	9,052,778	24,916,730	9,455,912	37.95%
E594	S25	iclC-1	0	5,991,675	7,940,724	7,155,052	19,251,804	12,949,495	67.26%
E595	S26	iclC-2	0	6,839,848	9,218,043	8,361,172	22,856,712	17,864,109	78.16%
E596	S27	iclC-3	0	7,254,089	10,014,821	9,636,648	25,176,694	15,836,513	62.90%
E597	S28	iclC-1	4	6,017,338	8,350,869	7,473,415	20,405,724	8,670,588	42.49%
E598	S29	iclC-2	4	6,546,781	9,009,950	8,748,220	22,560,512	9,900,506	43.88%
E599	S30	iclC-3	4	6,864,268	9,557,542	9,156,173	24,012,380	9,293,086	38.70%
E600	S31	iclC-1	8	7,327,088	9,855,745	8,902,573	24,320,636	19,309,180	79.39%
E601	S32	iclC-2	8	7,256,705	10,058,106	9,034,513	24,560,391	20,142,239	82.01%
E602	S33	iclC-3	8	7,423,799	10,226,270	9,043,798	25,081,167	8,072,448	32.19%
E603	S34	iclC-1	12	6,976,862	9,556,081	8,469,828	23,395,539	8,920,152	38.13%
E604	S35	iclC-2	12	7,780,492	10,572,880	9,398,663	25,931,665	19,266,856	74.30%
E607	S38	iclC-2	16	7,212,304	9,907,992	8,763,376	24,335,774	8,009,339	32.91%
E608	S39	iclC-3	16	7,725,427	10,495,602	9,337,691	25,786,338	15,740,171	61.04%
E609	S40	iclC-1	20	7,425,444	10,198,295	9,097,592	24,989,399	19,774,947	79.13%
E610	S41	iclC-2	20	7,582,576	10,550,976	9,460,866	25,936,035	20,327,813	78.38%
E611	S42	iclC-3	20	7,994,283	11,079,058	9,964,672	27,283,509	18,219,098	66.78%
E612	S43	iclC-1	24	8,570,183	11,827,426	10,616,979	29,080,836	26,421,509	90.86%
E613	S44	iclC-2	24	7,361,546	10,242,117	9,194,225	25,008,708	20,489,891	81.93%
E614	S45	iclC-3	24	7,427,254	10,310,636	9,284,406	25,280,845	13,996,128	55.36%
E615	S46	iclC-1	28	7,541,377	10,466,620	9,394,935	25,707,308	14,065,243	54.71%
E616	S47	iclC-2	28	7,941,287	11,178,096	9,952,481	27,369,675	13,964,405	51.02%
E617	S48	iclC-3	28	7,826,064	10,993,591	9,816,724	26,842,677	22,396,041	83.43%

Appendix 1. Reads information for each sample.

KEGG Pathway			Number of genes in our analysis	Total number of genes	KEGG Pathway			Number of genes in our analysis	Total number of genes
1	ko01200	Carbon metabolism	129	140	43	ko00196	Photosynthesis - antenna proteins	23	23
2	ko03010	Ribosome	123	125	44	ko00561	Glycerolipid metabolism	22	25
3	ko01230	Biosynthesis of amino acids	119	126	45	ko00900	Terpenoid backbone biosynthesis	22	23
4	ko03040	Spliceosome	107	110	46	ko00510	N-Glycan biosynthesis	21	22
5	ko00230	Purine metabolism	99	115	47	ko00400	Phenylalanine, tyrosine and tryptophan biosynthesis	21	21
6	ko03013	RNA transport	92	95	48	ko00920	Sulfur metabolism	21	21
7	ko04141	Protein processing in endoplasmic reticulum	77	82	49	ko00061	Fatty acid biosynthesis	20	22
8	ko00240	Pyrimidine metabolism	76	82	50	ko00280	Valine, leucine and isoleucine degradation	20	21
9	ko00190	Oxidative phosphorylation	66	69	51	ko00220	Arginine biosynthesis	20	21
10	ko03008	Ribosome biogenesis in eukaryotes	57	61	52	ko03022	Basal transcription factors	19	21
11	ko04120	Ubiquitin mediated proteolysis	55	64	53	ko00770	Pantothenate and CoA biosynthesis	18	20
12	ko03018	RNA degradation	54	61	54	ko03430	Mismatch repair	17	22
13	ko04144	Endocytosis	52	54	55	ko00640	Propanoate metabolism	17	19
14	ko04146	Peroxisome	45	50	56	ko00130	Ubiquinone and other terpenoid-quinone biosynthesis	17	19
15	ko03015	mRNA surveillance pathway	45	47	57	ko00760	Nicotinate and nicotinamide metabolism	17	19
16	ko00010	Glycolysis / Gluconeogenesis	44	49	58	ko00053	Ascorbate and aldarate metabolism	17	18
17	ko00620	Pyruvate metabolism	43	49	59	ko00071	Fatty acid degradation	17	17
18	ko00270	Cysteine and methionine metabolism	42	43	60	ko03410	Base excision repair	16	20
19	ko00970	Aminoacyl-tRNA biosynthesis	41	45	61	ko00600	Sphingolipid metabolism	16	18
20	ko00860	Porphyrin and chlorophyll metabolism	39	42	62	ko01040	Biosynthesis of unsaturated fatty acids	16	16
21	ko00500	Starch and sucrose metabolism	37	45	63	ko03440	Homologous recombination	15	22
22	ko00710	Carbon fixation in photosynthetic organisms	36	40	64	ko00350	Tyrosine metabolism	14	16
23	ko01212	Fatty acid metabolism	36	38	65	ko00450	Selenocompound metabolism	14	16
24	ko00520	Amino sugar and nucleotide sugar metabolism	35	37	66	ko04130	SNARE interactions in vesicular transport	14	15
25	ko04070	Phosphatidylinositol signaling system	35	36	67	ko00670	One carbon pool by folate	14	14
26	ko00260	Glycine, serine and threonine metabolism	34	39	68	ko00051	Fructose and mannose metabolism	13	19
27	ko00630	Glyoxylate and dicarboxylate metabolism	34	37	69	ko02010	ABC transporters	13	16
28	ko03050	Proteasome	34	34	70	ko00906	Carotenoid biosynthesis	13	15
29	ko00195	Photosynthesis	33	35	71	ko00340	Histidine metabolism	13	14
30	ko00330	Arginine and proline metabolism	32	36	72	ko00052	Galactose metabolism	12	16
31	ko00020	Citrate cycle (TCA cycle)	32	33	73	ko00590	Arachidonic acid metabolism	12	12
32	ko00480	Glutathione metabolism	32	32	74	ko00290	Valine, leucine and isoleucine biosynthesis	12	12
33	ko00564	Glycerophospholipid metabolism	31	38	75	ko04140	Regulation of autophagy	11	13
34	ko03420	Nucleotide excision repair	31	37	76	ko00360	Phenylalanine metabolism	11	12
35	ko01210	2-Oxocarboxylic acid metabolism	30	30	77	ko00100	Steroid biosynthesis	11	11
36	ko04145	Phagosome	29	30	78	ko04075	Plant hormone signal transduction	11	11

37	ko03060	Protein export	28	29	79	ko00410	beta-Alanine metabolism	10	12
38	ko03030	DNA replication	27	33	80	ko00592	alpha-Linolenic acid metabolism	10	11
39	ko00250	Alanine, aspartate and glutamate metabolism	26	27	81	ko00040	Pentose and glucuronate interconversions	10	11
40	ko03020	RNA polymerase	26	26	82	ko00300	Lysine biosynthesis	10	10
41	ko00030	Pentose phosphate pathway	24	26					
42	ko00562	Inositol phosphate metabolism	23	24	Total number of genes			2710	2937

693 **Appendix 2**. KEGG pathways defined for *C. reinhardtii* (http://www.kegg.jp/kegg/). Names of the pathways
694 (ko indexes), number of genes contained in the pathway and number of genes of the pathway detected in our
695 transcriptomic data are mentioned.

696

	balance state – all 16 samples together				
	Identifier	Name	G_{i0} value	Definition	Description
1	Cre03.g207050.v5.5	RPL29	-0.0441087	Ribosomal protein L29, component of cytosolic 80S ribosome and 60S large subunit	Cytosolic 80S ribosomal protein L29; Cytosolic 60S large ribosomal subunit protein L29
2	Cre10.g420750.v5.5	RPL30	-0.0417273	Ribosomal protein L3-0. component of cytosolic 80S ribosome and 60S large subunit	Cytosolic 80S ribosomal protein L30; Cytosolic 60S large ribosomal subunit protein L30
3	Cre12.g529400.v5.5	RPS27e 1	-0.0412797	Ribosomal protein S27e isoform 1, component of 80S ribosome and 40S small subunit	Cytosolic 80S ribosomal protein S27e, isoform 1; Cytosolic 40S small ribosomal subunit protein S27e, isoform 1. Previously annoated as RPS27-A
4	Cre16.g666301.v5.5	RPS30	-0.0401102	Ribosomal protein S3-0. component of cytosolic 80S ribosome and 40S small subunit	Cytosolic 80S ribosomal protein S30; Cytosolic 40S small ribosomal subunit protein S30
5	Cre06.g310700.v5.5	RPL36a	-0.0400216	Ribosomal protein L36a, component of cytosolic 80S ribosome and 60S large subunit	This gene has also been annotated as RPL41 (according to the nomenclature used for the yeast ribosomal subunits) and shown to be the site of the cycloheximide resistance mutation ACT2. See PMID: 11254126. Cytosolic 80S ribosomal protein L36a; Cytosolic
6	Cre03.g203450.v5.5	RPS21	-0.0394877	Ribosomal protein S21, component of cytosolic 80S ribosome and 40S small subunit	Cytosolic 80S ribosomal protein S21; Cytosolic 40S small ribosomal subunit protein S21
7	Cre06.g282500.v5.5	RPL23a	-0.0394199	Ribosomal protein L23a, component of cytosolic 80S ribosome and 60S large subunit	Cytosolic 80S ribosomal protein L23a; Cytosolic 60S large ribosomal subunit protein L23a
8	Cre08.g382500.v5.5	RPS25	-0.0394116	Ribosomal protein S25, component of cytosolic 80S ribosome and 40S small subunit	Cytosolic 80S ribosomal protein S25; Cytosolic 40S small ribosomal subunit protein S25
9	Cre08.g360900.v5.5	RPS15	-0.039343	Ribosomal protein S15, component of cytosolic 80S ribosome and 40S small subunit	Cytosolic 80S ribosomal protein S15; Cytosolic 40S small ribosomal subunit protein S15
10	Cre16.g682300.v5.5	RPS26	-0.0389401	Ribosomal protein S26, component of cytosolic 80S ribosome and 40S small subunit	Cytosolic 80S ribosomal protein S26; Cytosolic 40S small ribosomal subunit protein S26
11	Cre02.g143050.v5.5	RPP2	-0.0388384	Acidic ribosomal protein P2	Cytosolic 80S ribosomal protein P2; Cytosolic 60S large ribosomal subunit protein P2
12	Cre01.g066917.v5.5	LHCBM 1	-0.0383745	Chlorophyll a/b binding protein of LHCII	Chlorophyll a-b binding protein of LHCII
13	Cre02.g120150.v5.5	RBCS2	-0.0383725	Ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit 2	RuBisCO small subunit 2, chloroplast precursor [PMID: 3820291]
14	Cre12.g514500.v5.5	RPS11	-0.0383228	Ribosomal protein S11, component of cytosolic 80S ribosome and 40S small subunit	Cytosolic 80S ribosomal protein S11; Cytosolic 40S small ribosomal subunit protein S11
15	Cre12.g489153.v5.5	#N/A	-0.0381789	#N/A	#N/A
16	Cre07.g325746.v5.5	RPL38	-0.0380879	Ribosomal protein L38, component of cytosolic 80S ribosome and 60S large subunit	Cytosolic 80S ribosomal protein L38; Cytosolic 60S large ribosomal subunit protein L38
17	Cre08.g359750.v5.5	RPS9	-0.0379511	Ribosomal protein S9, component of cytosolic 80S ribosome and 40S small subunit	Cytosolic 80S ribosomal protein S9; Cytosolic 40S small ribosomal subunit protein S9
18	Cre10.g430400.v5.5	RPL37	-0.037616	Ribosomal protein L37, component of cytosolic 80S ribosome and 60S large subunit	Cytosolic 80S ribosomal protein L37; Cytosolic 60S large ribosomal subunit protein L37
19	Cre12.g494050.v5.5	RPL9	-0.037489	Ribosomal protein L9, component of cytosolic 80S ribosome and 60S large subunit	Cytosolic 80S ribosomal protein L9; Cytosolic 60S large ribosomal subunit protein L9
20	Cre09.g388200.v5.5	RPL10	-0.0373316	Ribosomal protein L1-0. component of cytosolic 80S ribosome and 60S large subunit	Cytosolic 80S ribosomal protein L10; Cytosolic 60S large ribosomal subunit protein L10
21	Cre10.g456200.v5.5	RPS24	-0.0371878	Ribosomal protein S24, component of cytosolic 80S ribosome and 40S small subunit	Cytosolic 80S ribosomal protein S24; Cytosolic 40S small ribosomal subunit protein S24. Possible locus for PR1 (paromomycin resistance) genetic marker
22	Cre04.g211800.v5.5	RPL23	-0.0371363	Ribosomal protein L23, component of cytosolic 80S ribosome and 60S	Cytosolic 80S ribosomal protein L23; Cytosolic 60S large ribosomal subunit protein L23

				large subunit	
23	**Cre12.g498250.v5.5**	**RPS17**	**-0.0370797**	**Ribosomal protein S17, component of cytosolic 80S ribosome and 40S small subunit**	**Cytosolic 80S ribosomal protein S17; Cytosolic 40S small ribosomal subunit protein S17**
24	Cre04.g214503.v5.5	#N/A	-0.0370682	#N/A	#N/A
24	**Cre06.g290950.v5.5**	**RPS5**	**-0.036391**	**Ribosomal protein S5, component of cytosolic 80S ribosome and 40S small subunit**	**Cytosolic 80S ribosomal protein S5; Cytosolic 40S small ribosomal subunit protein S5**
26	Cre03.g182551.v5.5	PCY1	-0.0362985	Pre-apoplastocyanin	Pre-apoplastocyanin copper binding protein, PETE [PMID: 2165059; PMID: 8940133]; structure of plastocyanin PDB: 2PLT; mutant = ac208 [PMID: 8463310]
27	**Cre12.g537800.v5.5**	**RPL7**	**-0.0360244**	**Ribosomal protein L7, component of cytosolic 80S ribosome and 60S large subunit**	**Cytosolic 80S ribosomal protein L7; Cytosolic 60S large ribosomal subunit protein L7**
28	Cre14.g626700.v5.5	PETF	-0.0360156	Ferredoxin	2Fe-2S containing redox protein involved in photosynthetic electron transfer, chloroplast localization [PMID: 16656453]
29	**Cre13.g568900.v5.5**	**RPL17**	**-0.0358546**	**Ribosomal protein L17, component of cytosolic 80S ribosome and 60S large subunit**	**Cytosolic 80S ribosomal protein L17; Cytosolic 60S large ribosomal subunit protein L17**
30	**Cre08.g358556.v5.5**	**RPS29**	**-0.0357983**	**Ribosomal protein S29, component of cytosolic 80S ribosome and 40S small subunit**	**Cytosolic 80S ribosomal protein S29; Cytosolic 40S small ribosomal subunit protein S29**
31	**Cre06.g272800.v5.5**	**RPS8**	**-0.0357831**	**Ribosomal protein S8, component of cytosolic 80S ribosome and 40S small subunit**	**Cytosolic 80S ribosomal protein S8; Cytosolic 40S small ribosomal subunit protein S8**
32	**Cre06.g273600.v5.5**	**RPS27a**	**-0.0357815**	**Ribosomal protein S27a, component of cytosolic 80S ribosome and 40S small subunit**	**Cytosolic 80S ribosomal protein S27a; Cytosolic 40S small ribosomal subunit protein S27a; ubiquitin region ends with amino acid residue 74. Has homology to ubiquitin superfamily**
33	**Cre12.g512600.v5.5**	**RPL18**	**-0.0356463**	**Ribosomal protein L18, component of cytosolic 80S ribosome and 60S large subunit**	**Cytosolic 80S ribosomal protein L18; Cytosolic 60S large ribosomal subunit protein L18**
34	**Cre10.g459250.v5.5**	**RPL35a**	**-0.0354754**	**Ribosomal protein L35a, component of cytosolic 80S ribosome and 60S large subunit**	**Cytosolic 80S ribosomal protein L35a; Cytosolic 60S large ribosomal subunit protein L35a**
35	**Cre17.g738300.v5.5**	**RPP1**	**-0.0354434**	**Acidic ribosomal protein P1**	**Cytosolic 80S ribosomal protein P1; Cytosolic 60S large ribosomal subunit protein P1**
36	Cre10.g420350.v5.5	PSAE	-0.0353805	Photosystem I 8.1 kDa reaction center subunit IV	Photosystem I reaction center subunit IV, chloroplast precursor (PSI-E) (Photosystem I 8.1 kDa protein) [PMID: 2693938]
37	**Cre01.g027000.v5.5**	**RPL11**	**-0.0353561**	**Ribosomal protein L11, component of cytosolic 80S ribosome and 60S large subunit**	**Cytosolic 80S ribosomal protein L11; Cytosolic 60S large ribosomal subunit protein L11**
38	**Cre17.g701650.v5.5**	**RPL27**	**-0.0352832**	**Ribosomal protein L27, component of cytosolic 80S ribosome and 60S large subunit**	**Cytosolic 80S ribosomal protein L27; Cytosolic 60S large ribosomal subunit protein L27**
39	**Cre14.g617900.v5.5**	**RPL35**	**-0.0352725**	**Ribosomal protein L35, component of cytosolic 80S ribosome and 60S large subunit**	**Cytosolic 80S ribosomal protein L35; Cytosolic 60S large ribosomal subunit protein L35**
40	**Cre02.g106600.v5.5**	**RPS19**	**-0.0352276**	**Ribosomal protein S19, component of cytosolic 80S ribosome and 40S small subunit**	**Cytosolic 80S ribosomal protein S19; Cytosolic 40S small ribosomal subunit protein S19**
41	**Cre12.g532550.v5.5**	**RPL13a**	**-0.0352177**	**Ribosomal protein L13a, component of cytosolic 80S ribosome and 60S large subunit**	**Cytosolic 80S ribosomal protein L13a; Cytosolic 60S large ribosomal subunit protein L13a**
42	**Cre02.g102250.v5.5**	**RPS3**	**-0.0351398**	**Ribosomal protein S3, component of cytosolic 80S ribosome and 40S small subunit**	**Cytosolic 80S ribosomal protein S3; Cytosolic 40S small ribosomal subunit protein S3**
43	**Cre06.g257150.v5.5**	**RPL37a**	**-0.035103**	**Ribosomal protein L37a, component of cytosolic 80S ribosome and 60S large subunit**	**Cytosolic 80S ribosomal protein L37a; Cytosolic 60S large ribosomal subunit protein L37a**
44	**Cre12.g498900.v5.5**	**RPS7**	**-0.0350799**	**Ribosomal protein S7, component of cytosolic 80S ribosome and 40S small subunit**	**Cytosolic 80S ribosomal protein S7; Cytosolic 40S small ribosomal subunit protein S7**
45	**Cre12.g528750.v5.5**	**RPL12**	**-0.0350349**	**Ribosomal protein L12, component of cytosolic 80S ribosome and 60S large subunit**	**Cytosolic 80S ribosomal protein L12; Cytosolic 60S large ribosomal subunit protein L12**
46	**Cre07.g357850.v5.5**	**RPL22**	**-0.034811**	**Ribosomal protein L22, component of cytosolic 80S ribosome and 60S large subunit**	**Cytosolic 80S ribosomal protein L22; Cytosolic 60S large ribosomal subunit protein L22**

47	**Cre12.g504200.v5.5**	**RPS23**	**-0.0344687**	**Ribosomal protein S23, component of cytosolic 80S ribosome and 40S small subunit**	**Cytosolic 80S ribosomal protein S23; Cytosolic 40S small ribosomal subunit protein S23**
48	Cre09.g405106.v5.5	#N/A	-0.0343005	#N/A	#N/A
49	**Cre12.g484050.v5.5**	**RPL36**	**-0.0342738**	**Ribosomal protein L36, component of cytosolic 80S ribosome and 60S large subunit**	**Cytosolic 80S ribosomal protein L36; Cytosolic 60S large ribosomal subunit protein L36**
50	**Cre09.g391097.v5.5**	**RPL24**	**-0.0342292**	**Ribosomal protein L24, component of cytosolic 80S ribosome and 60S large subunit**	**Cytosolic 80S ribosomal protein L24; Cytosolic 60S large ribosomal subunit protein L24**
51	Cre01.g007051.v5.5	#N/A	-0.0341731	#N/A	#N/A
52	Cre09.g402219.v5.5	LCI3	-0.0341695	Low-CO2-inducible protein	Regulated by CCM1 [PMID: 15235119]. Acclimation to changing CO2 concentrations and light intensities was studied by Yamano et al. 2008 [PMID: 18322145].
53	**Cre14.g621450.v5.5**	**RPL5**	**-0.0337335**	**Ribosomal protein L5, component of cytosolic 80S ribosome and 60S large subunit**	**Cytosolic 80S ribosomal protein L5; Cytosolic 60S large ribosomal subunit protein L5**
54	Cre12.g498600.v5.5	#N/A	-0.0337325	Eukaryotic translation elongation factor 1 alpha	Flagellar Associated Protein, found in the flagellar proteome [PMID: 15998802]. Previously annotated as just EEF1
55	**Cre17.g701200.v5.5**	**RPL14**	**-0.0336314**	**Ribosomal protein L14, component of cytosolic 80S ribosome and 60S large subunit**	**Cytosolic 80S ribosomal protein L14; Cytosolic 60S large ribosomal subunit protein L14**
56	**Cre13.g568650.v5.5**	**RPS3a**	**-0.0334977**	**Ribosomal protein S3a, component of cytosolic 80S ribosome and 40S small subunit**	**Cytosolic 80S ribosomal protein S3a; Cytosolic 40S small ribosomal subunit protein S3a**
57	**Cre01.g047750.v5.5**	**RPL18a**	**-0.0334572**	**Ribosomal protein L18a, component of cytosolic 80S ribosome and 60S large subunit**	**Cytosolic 80S ribosomal protein L18a; Cytosolic 60S large ribosomal subunit protein L18a**
58	**Cre06.g278135.v5.5**	**RPL21**	**-0.0331582**	**Ribosomal protein L21, component of cytosolic 80S ribosome and 60S large subunit**	**Cytosolic 80S ribosomal protein L21; Cytosolic 60S large ribosomal subunit protein L21**
59	**Cre02.g101350.v5.5**	**RPL10a**	**-0.0331156**	**Ribosomal protein L10a, component of cytosolic 80S ribosome and 60S large subunit**	**Cytosolic 80S ribosomal protein L10a; Cytosolic 60S large ribosomal subunit protein L10a**
60	**Cre02.g115200.v5.5**	**RPL27a**	**-0.033038**	**Ribosomal protein L27a, component of cytosolic 80S ribosome and 60S large subunit**	**Cytosolic 80S ribosomal protein L27a; Cytosolic 60S large ribosomal subunit protein L27a. Candidate gene for the cycloheximide resistance mutation act1.**
61	Cre12.g546150.v5.5	PETM	-0.0329136	Cytochrome b6f complex PetM subunit	Cytochrome b6f complex chain PetM, chloroplast precursor; GI:2493687; PMID: 8631873, PMID: 8616155, PMID: 7493968
62	Cre06.g278222.v5.5	RCK1	-0.0327044	Receptor of activated protein kinase C	Receptor of activated protein kinase C 1, component of 40S small ribosomal subunit; Also cytosolic 40S small ribosomal subunit protein RACK1. Previously annotated as RACK1 and CBLP. Initially described [PMID: 2116589] as CBLP. Smith and Lee 2008 [PMID:
63	Cre12.g560950.v5.5	PSAG	-0.0325305	Photosystem I reaction center subunit V	(PSI-G) (P35 protein) [PMID: 2693938]
64	**Cre10.g417700.v5.5**	**RPL3**	**-0.0325262**	**Ribosomal protein L3, component of cytosolic 80S ribosome and 60S large subunit**	**Cytosolic 80S ribosomal protein L3; Cytosolic 60S large ribosomal subunit protein L3**
65	**Cre02.g075700.v5.5**	**RPL19**	**-0.0325095**	**Ribosomal protein L19, component of cytosolic 80S ribosome and 60S large subunit**	**Cytosolic 80S ribosomal protein L19; Cytosolic 60S large ribosomal subunit protein L19**
66	Cre12.g486300.v5.5	PSAL	-0.0324252	Photosystem I reaction center subunit XI	#N/A
67	Cre12.g529651.v5.5	#N/A	-0.0324189	#N/A	#N/A
68	Cre11.g467578.v5.5	#N/A	-0.0322713	#N/A	#N/A
69	**Cre07.g331900.v5.5**	**RPS13**	**-0.032224**	**Ribosomal protein S13, component of cytosolic 80S ribosome and 40S small subunit**	**Cytosolic 80S ribosomal protein S13; Cytosolic 40S small ribosomal subunit protein S13**
70	**Cre06.g272950.v5.5**	**RPS18**	**-0.0321851**	**Ribosomal protein S18, component of cytosolic 80S ribosome and 40S small subunit**	**Cytosolic 80S ribosomal protein S18; Cytosolic 40S small ribosomal subunit protein S18**
71	**Cre09.g411100.v5.5**	**RPS10**	**-0.032052**	**Ribosomal protein S1-0. component of cytosolic 80S ribosome and 40S small subunit**	**Cytosolic 80S ribosomal protein S10; Cytosolic 40S small ribosomal subunit protein S10**
72	**Cre14.g630100.v5.5**	**RPL13**	**-0.0319474**	**Ribosomal protein L13, component of cytosolic 80S ribosome and 60S large subunit**	**Cytosolic 80S ribosomal protein L13; Cytosolic 60S large ribosomal subunit protein L13**
73	Cre13.g577100.v5.5	ACP2	-0.0317649	Acyl-carrier protein	Acyl-carrier protein

74	Cre12.g483850.v5.5	#N/A	-0.0317034	#N/A	#N/A
75	Cre06.g263450.v5.5	#N/A	-0.0315772	#N/A	#N/A
76	**Cre02.g091100.v5.5**	**RPL15**	**-0.031471**	**Ribosomal protein L15, component of cytosolic 80S ribosome and 60S large subunit**	**Cytosolic 80S ribosomal protein L15; Cytosolic 60S large subunit ribosomal protein L15**
77	**Cre01.g011000.v5.5**	**RPL6**	**-0.0314481**	**Ribosomal protein L6, component of cytosolic 80S ribosome and 60S large subunit**	**Cytosolic 80S ribosomal protein L6; Cytosolic 60S large ribosomal subunit protein L6**
78	Cre05.g234637.v5.5	#N/A	-0.031324	#N/A	#N/A
79	**Cre12.g520500.v5.5**	**RPP0**	**-0.0312926**	**Acidic ribosomal protein P-0. Ribosomal protein L10**	**Cytosolic 80S acidic ribosomal protein P0; Cytosolic 60S large ribosomal subunit protein P0. Ribosomal protein L10**
80	Cre12.g548950.v5.5	LHCBM7	-0.0309816	Chlorophyll a/b binding protein of LHCII	Chlorophyll a-b binding protein of LHCII
81	**Cre06.g249250.v5.5**	**RPL7ae**	**-0.0309195**	**Ribosomal protein L7Ae**	**Cytosolic 80S ribosomal protein L7ae; Cytosolic 60S large ribosomal subunit protein L7ae**
82	**Cre16.g661050.v5.5**	**RPL34**	**-0.0307426**	**Ribosomal protein, L34e superfamily, component of cytosolic 80S ribosome and 60S large subunit**	**Cytosolic 80S ribosomal protein L34; Cytosolic 60S large ribosomal subunit protein L34. Belongs to Ribosomal_L34e superfamily**
83	Cre12.g548400.v5.5	LHCBM2	-0.0307266	Light-harvesting protein of photosystem II	Encoding a light-harvesting antenna protein for PS2. This gene was reported under the name of LhcII-3 in PMID: 11522911, and the sequence has been deposited in Genbank (AB051205). Also designated as LhcbM2 in PMID: 16143838 and as Lhcbm2 in PMID: 14652691
84	Cre12.g535851.v5.5	#N/A	-0.0306301	#N/A	#N/A
85	Cre16.g660150.v5.5	#N/A	-0.0306154	#N/A	#N/A
86	**Cre01.g040000.v5.5**	**RPL26**	**-0.0306063**	**Ribosomal protein L26, component of cytosolic 80S ribosome and 60S large subunit**	**Cytosolic 80S ribosomal protein L26; Cytosolic 60S large ribosomal subunit protein L26**
87	**Cre01.g039250.v5.5**	**RPS2**	**-0.0305367**	**Ribosomal protein S2, component of cytosolic 80S ribosome and 40S small subunit**	**Cytosolic 80S ribosomal protein S2; Cytosolic 40S small ribosomal subunit protein S2**
88	**Cre09.g397697.v5.5**	**RPL4**	**-0.0302693**	**Ribosomal protein L4, component of cytosolic 80S ribosome and 60S large subunit**	**Cytosolic 80S ribosomal protein L4; Cytosolic 60S large ribosomal subunit protein L4**
89	Cre06.g285250.v5.5	LHCBM6	-0.0301451	Chloropyll a/b binding protein of LHCII type I, chloroplast precursor	Chloropyll a-b binding protein of LHCII type I, chloroplast precursor
90	Cre17.g724300.v5.5	PSAK	-0.0301443	Photosystem I reaction center subunit psaK	8.4 kD subunit of photosystem I (polypeptide 37) [PMID: 2693938]
91	**Cre12.g510450.v5.5**	**RPS28**	**-0.0301316**	**Ribosomal protein S28, component of cytosolic 80S ribosome and 40S small subunit**	**Cytosolic 80S ribosomal protein S28; Cytosolic 40S small ribosomal subunit protein S28**
92	**Cre06.g308250.v5.5**	**RPS4**	**-0.0301189**	**Ribosomal protein S4, component of cytosolic 80S ribosome and 40S small subunit**	**Cytosolic 80S ribosomal protein S4; Cytosolic 40S small ribosomal subunit protein S4**
93	Cre06.g272650.v5.5	LHCA8	-0.0300836	Light-harvesting protein of photosystem I	#N/A
94	**Cre09.g400650.v5.5**	**RPS6**	**-0.0299611**	**Ribosomal protein S6, component of cytosolic 80S ribosome and 40S small subunit**	**Cytosolic 80S ribosomal protein S6; Cytosolic 40S small ribosomal subunit protein S6**
95	Cre01.g010900.v5.5	GAP3	-0.0299284	Glyceraldehyde-3-Phosphate Dehydrogenase	Glyceraldehyde 3-phosphate dehydrogenase A, chloroplast precursor (NADP-dependent glyceraldehyde phosphate dehydrogenase subunit A); corresponds to G3PA_CHLRE; found in the flagellar proteome [PMID: 15998802]. This enzyme is bispecific for NADP and NAD,
96	**Cre17.g734450.v5.5**	**PRPL19**	**-0.029736**	**Plastid ribosomal protein L19**	**Chloroplast ribosomal protein L19, imported to chloroplast; Chloroplast large ribosomal subunit protein L19**
97	Cre06.g261000.v5.5	PSBR	-0.0296755	10 kDa photosystem II polypeptide	Similar to At1g7904-0. chloroplast-targeted
98	Cre04.g214150.v5.5	THI4	-0.0296257	Thiazole biosynthetic enzyme; THI4 regulatory protein	Involved in the synthesis of the thiazole moiety of thiamine pyrophosphate. One of 2 splice variants.; Alternatively spliced variant of THI4 involved in the regulation of THI4a. Contains riboswitch THI RNA motif at positions 426525-426691. Regulates p
99	Cre17.g702950.v5.5	#N/A	-0.0295457	#N/A	#N/A
100	Cre16.g650550.v5.5	FAP103	-0.0294053	Flagellar Associated Protein, nucleoside diphosphate kinase-like	Flagellar Associated Protein similar to nucleoside diphosphate kinase, found in the flagellar proteome [PMID: 15998802]

Appendix 3. List of the 100 most highly expressed genes in the balance constraint when all 16 samples are grouped together. *C. reinhardtii* genome version v5.5. #N/A means the function of the gene is unknown.

	Light samples – constraint 1 – 100 most positive genes- *icIC* phenotype				
	Identifier	Name	G_{i1} value	Definition	Description
1	Cre02.g141206.v5.5	#N/A	0.061958908	#N/A	#N/A
2	Cre13.g577850.v5.5	#N/A	0.049663419	Peptidyl-prolyl cis-trans isomerase, FKBP-type	FKBP-type peptidyl-prolyl cis-trans isomerase (EC 5.2.1.8) (PPIase) (Rotamase); probably targeted to thylakoid lumen (TP length 31 + RR motif); [PMID: 15701785]
3	Cre03.g179820.v5.5	#N/A	0.048585519	#N/A	#N/A
4	Cre16.g659300.v5.5	#N/A	0.047300183	Cytochrome b5 protein	Cytochrome b5 protein with possible nitrate reductase activity. Previously annotated as CYB5-2
5	Cre03.g144807.v5.5	MAS1	0.045825161	Malate synthase	Malate synthase (EC 2.3.3.9); identical to cDNA sequence (AAP75564); PMID: 19214701
6	Cre12.g546550.v5.5	FEA1	0.040314599	Fe-assimilating protein	High-CO2 inducible, iron-deficiency inducible, periplasmic protein. Changes in expression with light intensity were studied by Long and Merchant 2008 [PMID: 19067961]. Cis-acting regulatory elements were characterized by Fei et al. 2009 [PMID: 19351705].
7	Cre06.g282800.v5.5	ICL1	0.039174799	Isocitrate lyase	Isocitrate lyase (EC 4.1.3.1); isocitrase; 98% identical to cDNA (AAB61446) [PMID: 9049260]; PMID: 19847013; Listed by Rolland et al. 2008 [PMID: 19451016] in a table of putative redox-regulated proteins identified by proteomic methods
8	Cre17.g702900.v5.5	#N/A	0.038402915	#N/A	#N/A
9	Cre01.g032650.v5.5	TAL1	0.035335734	Transaldolase	Putative Transaldolase (EC 2.2.1.2); possible cytosolic isoform, although weak predicted organelle targeting by Target-P; catalyzes the reversible transfer of a three-carbon ketol unit from sedoheptulose 7-phosphate to glyceraldehyde 3-phosphate to form e
10	Cre08.g364100.v5.5	#N/A	0.029068053	#N/A	#N/A
11	Cre12.g550500.v5.5	#N/A	0.028717883	#N/A	#N/A
12	Cre07.g353450.v5.5	ACS3	0.028705257	Acetyl-CoA synthetase/ligase	Acetyl-CoA synthetase (EC 6.2.1.1); Acetate-CoA ligase; probable mitochondrial protein based on mass spectrometry identification (QFYTAPTLLR + SLLQLGDAWPR), although organelle targeting predicted as other by Target-P and iPSORT. AMP-dependent
13	Cre14.g630000.v5.5	KIR2	0.028376081	Ketoacid isomerase-like protein	Related to a series of bacterial genes among which is a putative ketosteroid isomerase-related protein of Ralstonia; however, it is impossible to find a similarity with the well-characterized delta5-3-ketosteroid isomerase of Pseudomonas putida or Comamon
14	Cre16.g669600.v5.5	#N/A	0.027608332	#N/A	#N/A
15	Cre10.g428100.v5.5	#N/A	0.027595198	#N/A	#N/A
16	Cre02.g141166.v5.5	#N/A	0.026776304	#N/A	#N/A
17	Cre12.g540500.v5.5	#N/A	0.024650013	#N/A	#N/A
18	Cre02.g094250.v5.5	#N/A	0.024275881	#N/A	#N/A
19	Cre14.g623176.v5.5	#N/A	0.023773093	#N/A	#N/A
20	Cre12.g523900.v5.5	#N/A	0.02374743	#N/A	#N/A
21	Cre06.g263300.v5.5	#N/A	0.023687946	#N/A	#N/A
22	Cre02.g141400.v5.5	PCK1	0.023213075	Phosphoenolpyruvate carboxykinase	Phosphoenolpyruvate carboxykinase; PEP carboxykinase (EC 4.1.1.49); based on high similarity to PEPCK from Panicum maximum (GenBank AAQ10076) and many other plants. Two splice variants are expected, which may result in cytoplasmic and plastid-targeted fo
23	Cre01.g044450.v5.5	#N/A	0.023071184	#N/A	#N/A
24	Cre12.g543400.v5.5	FDH1	0.022667243	Formaldehyde dehydrogenase	Formaldehyde dehydrogenase, glutathione-dependent, zinc-containing alcohol dehydrogenase class III; found in the flagellar proteome [PMID: 15998802]
24	Cre09.g387726.v5.5	AST1	0.022344824	Aspartate aminotransferase	Putative aspartate aminotransferase (EC 2.6.1.1); no organelle targeting sequence predicted by Target-P, but plastidic by homology, Found in the mitochondrial

					proteomic survey
26	Cre13.g573000.v5.5	#N/A	0.02193045	#N/A	#N/A
27	Cre07.g315400.v5.5	#N/A	0.021746024	Simlar to nitric oxide associated protein protein of CSG family	Putative GTPase; shows similarity to plant nitric oxide associated protein (NOA1)
28	Cre12.g541250.v5.5	NAR1.5	0.02147352	Nitrite transporter	Putative nitrite transporter, possibly chloroplastic [PMID: 11912227; accession AY612641]
29	Cre15.g641200.v5.5	#N/A	0.021410352	#N/A	#N/A
30	Cre08.g358558.v5.5	#N/A	0.021337542	#N/A	#N/A
31	Cre50.g761397.v5.5	#N/A	0.021255316	#N/A	#N/A
32	Cre03.g149100.v5.5	CIS2	0.021172986	Citrate synthase	Citrate synthase (EC 2.3.3.1), glyoxysomal/microbody form; similarity to Arabidopsis citrate synthase glyoxysomal precursor (GenBank Q9LXS6); PMID: 19214701
33	Cre04.g214500.v5.5	IDH3	0.020938003	Isocitrate dehydrogenase, NADP-dependent	NADP specific isocitrate dehydrogenase (EC 1.1.1.42), mitochondrial precursor; similar to mammalian mitochondrial ICDH (e.g., GenBank XP_536192), predicted by Target-P to have an organellar targeting sequence (mitochondrial)
34	Cre17.g731900.v5.5	#N/A	0.020728269	#N/A	#N/A
35	Cre02.g108350.v5.5	#N/A	0.020164334	#N/A	#N/A
36	Cre17.g702950.v5.5	#N/A	0.019631584	#N/A	#N/A
37	Cre07.g332800.v5.5	#N/A	0.01928964	#N/A	#N/A
38	Cre14.g626000.v5.5	#N/A	0.019165025	#N/A	#N/A
39	Cre03.g153400.v5.5	#N/A	0.018845962	#N/A	#N/A
40	Cre13.g563500.v5.5	#N/A	0.018756131	#N/A	#N/A
41	Cre07.g346650.v5.5	#N/A	0.0184522	#N/A	#N/A
42	Cre07.g328075.v5.5	#N/A	0.01828606	#N/A	#N/A
43	Cre03.g198850.v5.5	#N/A	0.018140209	#N/A	#N/A
44	Cre03.g153450.v5.5	#N/A	0.018020779	#N/A	#N/A
45	Cre06.g278136.v5.5	#N/A	0.017667186	#N/A	#N/A
46	Cre12.g523550.v5.5	#N/A	0.017591383	#N/A	#N/A
47	Cre01.g003200.v5.5	#N/A	0.017314642	#N/A	#N/A
48	Cre09.g411300.v5.5	#N/A	0.017221772	#N/A	#N/A
49	Cre03.g158100.v5.5	#N/A	0.016794902	#N/A	#N/A
50	Cre12.g541200.v5.5	#N/A	0.016733218	#N/A	#N/A
51	Cre08.g358553.v5.5	#N/A	0.016726931	#N/A	#N/A
52	Cre03.g164550.v5.5	#N/A	0.016485771	#N/A	#N/A
53	Cre01.g022283.v5.5	#N/A	0.016398348	#N/A	#N/A
54	Cre13.g591400.v5.5	#N/A	0.016366362	#N/A	#N/A
55	Cre13.g603750.v5.5	#N/A	0.016192035	#N/A	#N/A
56	Cre17.g705450.v5.5	LCI26	0.016098792	Low-CO2-induced U-box protein	Contains Ubox (smart00504), Modified RING finger domain; Modified RING finger domain, without the full complement of Zn2+-binding ligands. Probable involvement in E2-dependent ubiquitination. SOSUI prediction: 5 transmembrane region. Low-CO2 inducible gen
57	Cre01.g011630.v5.5	#N/A	0.01602515	#N/A	#N/A
58	Cre01.g043450.v5.5	#N/A	0.015978266	#N/A	#N/A
59	Cre07.g335700.v5.5	#N/A	0.015634359	#N/A	#N/A
60	Cre16.g679781.v5.5	#N/A	0.01547045	#N/A	#N/A
61	Cre06.g278215.v5.5	#N/A	0.015262739	#N/A	#N/A
62	Cre11.g479100.v5.5	#N/A	0.015122331	#N/A	#N/A
63	Cre13.g562300.v5.5	#N/A	0.015107314	#N/A	#N/A
64	Cre04.g215050.v5.5	#N/A	0.015000459	Beta-carotene hydroxylase, putative chloroplast precursor	Beta-carotene hydroxylase (carotene beta-hydroxylase), putative chloroplast precursor (CBH) [PMID:15849308]. Previously annotated as CHYB
65	Cre10.g421800.v5.5	#N/A	0.014944666	#N/A	#N/A
66	Cre02.g076300.v5.5	UPD2	0.014900528	Uroporphyrinogen decarboxylase	Uroporphyrinogen decarboxylase, chloroplast precursor (URO-D) (UPD) (HemE). Previously annotated as UROD2
67	Cre06.g278147.v5.5	#N/A	0.014727216	#N/A	#N/A
68	Cre09.g393150.v5.5	FOX1	0.014716875	Multicopper ferroxidase	Multicopper ferroxidase, iron-deficiency inducible; plasma membrane localized [PMID: 12481087]. Expression of this gene under iron-deficient conditions was studied by Busch et al. 2008 [PMID: 18363784]. Cis-acting regulatory elements were characterized b

69	Cre03.g167250.v5.5	#N/A	0.014651038	#N/A	#N/A
70	Cre12.g495900.v5.5	#N/A	0.014423644	#N/A	#N/A
71	Cre08.g384700.v5.5	#N/A	0.014376529	#N/A	#N/A
72	Cre02.g116250.v5.5	#N/A	0.01435141	#N/A	#N/A
73	Cre06.g265900.v5.5	#N/A	0.014285923	#N/A	#N/A
74	Cre14.g630859.v5.5	#N/A	0.014228269	#N/A	#N/A
75	Cre09.g391400.v5.5	#N/A	0.014133241	#N/A	#N/A
76	Cre12.g487150.v5.5	#N/A	0.014055305	#N/A	#N/A
77	Cre02.g119000.v5.5	#N/A	0.014049176	#N/A	#N/A
78	Cre05.g241655.v5.5	#N/A	0.01403316	#N/A	#N/A
79	Cre14.g633903.v5.5	#N/A	0.014025017	#N/A	#N/A
80	Cre19.g751047.v5.5	#N/A	0.01396973	#N/A	#N/A
81	Cre06.g278148.v5.5	#N/A	0.013866235	#N/A	#N/A
82	Cre12.g495850.v5.5	#N/A	0.013846479	#N/A	#N/A
83	Cre02.g109600.v5.5	#N/A	0.013595351	Inositol monophosphatase	Inositol monophosphatase; 93% identical to putative IMPase from Chlamydomonas incerta (GenBank AAW82030). Possible glutathione S-transferase activity. Previously annotated as INM1
84	Cre14.g630847.v5.5	#N/A	0.013594571	#N/A	#N/A
85	Cre17.g700750.v5.5	#N/A	0.013589104	#N/A	#N/A
86	Cre14.g615450.v5.5	#N/A	0.013579504	#N/A	#N/A
87	Cre07.g346600.v5.5	#N/A	0.01337419	#N/A	#N/A
88	Cre50.g761447.v5.5	#N/A	0.013362153	#N/A	#N/A
89	Cre17.g718150.v5.5	#N/A	0.013351936	#N/A	#N/A
90	Cre17.g700950.v5.5	FDX5	0.01328312	Apoferredoxin	Fe2S2 containing redox protein, predicted chloroplast localization
91	Cre17.g703700.v5.5	SCL2	0.013280228	Succinate-CoA ligase beta chain	Succinate-CoA ligase beta chain, mitochondrial precursor; Similar to Arabidopsis mitochondrial SCL (GenBank NP_179632). Two splice variants are expected . Previously annotated as SCLB1
92	Cre06.g254150.v5.5	UTP1	0.01323069	Nucleolar protein, component of the U3 processome	Associates with U3 snoRNA, required for pre-rRNA processing; Conserved 90S pre-ribosomal component essential for endonucleolytic cleavage of the 35 S rRNA precursor at A0. A1, and A2 sites; contains WD-40 repeats. Independent cDNA sequence: Genbank AY5963
93	Cre12.g547650.v5.5	#N/A	0.013198418	#N/A	#N/A
94	Cre12.g548950.v5.5	LHCBM7	0.01307735	Chlorophyll a/b binding protein of LHCII	Chlorophyll a-b binding protein of LHCII
95	Cre02.g073850.v5.5	CGL54	0.012997358	Predicted protein	Conserved expressed protein of unknown function, related to Photosystem II 11 kDa protein precursor
96	Cre01.g020305.v5.5	#N/A	0.012979289	#N/A	#N/A
97	Cre04.g232104.v5.5	LHCBM3	0.01295075	Light-harvesting complex II chlorophyll a/b binding protein M3	#N/A
98	Cre03.g182100.v5.5	#N/A	0.012872499	#N/A	#N/A
99	Cre09.g387875.v5.5	IPY3	0.012857272	Soluble inorganic pyrophosphatase	Soluble inorganic pyrophosphatase; corresponds to CAC42763.1 (ppaII)
100	Cre07.g329767.v5.5	#N/A	0.012845994	#N/A	#N/A

Appendix 4. List of the 100 most positive expressed genes (*iclC* phenotype) of the first constraint when 8 light samples are analysed separately. *C. reinhardtii* genome version v5.5. #N/A means the function of the gene is unknown.

Dark samples – constraint 2 – 100 most positive genes- *icIC* phenotype					
	Identifier	Name	G_{i2} value	Definition	Description
1	Cre02.g141206.v5.5	#N/A	0.05235323	#N/A	#N/A
2	Cre17.g700950.v5.5	FDX5	0.046540514	Apoferredoxin	Fe2S2 containing redox protein, predicted chloroplast localization
3	Cre02.g141166.v5.5	#N/A	0.043979616	#N/A	#N/A
4	Cre06.g282800.v5.5	ICL1	0.038810693	Isocitrate lyase	Isocitrate lyase (EC 4.1.3.1); isocitrase; 98% identical to cDNA (AAB61446) [PMID: 9049260]; PMID: 19847013; Listed by Rolland et al. 2008 [PMID: 19451016] in a table of putative redox-regulated proteins identified by proteomic methods
5	Cre01.g011630.v5.5	#N/A	0.036308776	#N/A	#N/A
6	Cre12.g546550.v5.5	FEA1	0.036183813	Fe-assimilating protein	High-CO2 inducible, iron-deficiency inducible, periplasmic protein. Changes in expression with light intensity were studied by Long and Merchant 2008 [PMID: 19067961]. Cis-acting regulatory elements were characterized by Fei et al. 2009 [PMID: 19351705].
7	Cre16.g659300.v5.5	#N/A	0.031419697	Cytochrome b5 protein	Cytochrome b5 protein with possible nitrate reductase activity. Previously annotated as CYB5-2
8	Cre03.g180750.v5.5	MES1	0.031153203	Cobalamin-independent methionine synthase	5-methyltetrahydropteroyltriglutamate-homocysteine S-methyltransferase (EC 2.1.1.14); Vitamin B12 (cobalamin) independent methionine synthase; homocysteine methylase; methyltetrahydropteroylpolyglutamate:homocysteine methyltransferase; (HMT); found in the
9	Cre08.g364100.v5.5	#N/A	0.031011559	#N/A	#N/A
10	Cre12.g546600.v5.5	FEA2	0.027133959	Fe-assimilating protein	High-CO2 inducible, periplasmic protein. Cis-acting regulatory elements were characterized by Fei et al. 2009 [PMID: 19351705].
11	Cre01.g036950.v5.5	#N/A	0.027005811	#N/A	#N/A
12	Cre10.g417600.v5.5	#N/A	0.025765966	#N/A	#N/A
13	Cre09.g389356.v5.5	#N/A	0.025658813	#N/A	#N/A
14	Cre12.g543400.v5.5	FDH1	0.025095061	Formaldehyde dehydrogenase	Formaldehyde dehydrogenase, glutathione-dependent, zinc-containing alcohol dehydrogenase class III; found in the flagellar proteome [PMID: 15998802]
15	Cre07.g356600.v5.5	#N/A	0.02498279	#N/A	#N/A
16	Cre12.g541400.v5.5	#N/A	0.024197881	#N/A	#N/A
17	Cre13.g586300.v5.5	FKB12	0.023605483	Peptidyl-prolyl cis-trans isomerase, FKBP-type	FKBP-type peptidyl-prolyl cis-trans isomerase (EC 5.2.1.8) (PPiase) (Rotamase); probably cytosolic (by homology to FKB1_HUMAN and AtFBKP12); possibly involved in signaling, in conjunction with the TOR (target of Rapamycin) protein kinase [PMID: 15270681]
18	Cre12.g555500.v5.5	#N/A	0.023196984	#N/A	#N/A
19	**Cre01.g062172.v5.5**	**HBV1**	**0.022852115**	**Histone H2B variant**	**Histone H2B variant. ChromoDB HTB3404**
20	Cre07.g344400.v5.5	#N/A	0.022680591	#N/A	#N/A
21	Cre02.g080450.v5.5	#N/A	0.022227943	#N/A	#N/A
22	Cre03.g144967.v5.5	#N/A	0.022135392	#N/A	#N/A
23	Cre06.g293950.v5.5	SHMT2	0.021907016	Serine hydroxymethyltransferase 2	Serine hydroxymethyltransferase, by homology to Arabidopsis gene (Identifier 023254); found in the flagellar proteome [PMID: 15998802]
24	Cre14.g632100.v5.5	#N/A	0.021716931	#N/A	#N/A
24	Cre12.g520050.v5.5	#N/A	0.021493012	#N/A	#N/A
26	**Cre09.g399911.v5.5**	**CDC20**	**0.021230101**	**Activator and specificity subunit of anaphase promoting complex**	**CDC20 homolog. Activator and specificity subunit for anaphase promoting complex. Subunit 1 of Anaphase Promoting Complex. Cell cycle regulated E3 ubiquitin ligase. Related to fizzy/cdh1 proteins**
27	Cre14.g619450.v5.5	#N/A	0.021116712	#N/A	#N/A
28	Cre13.g586000.v5.5	#N/A	0.020674421	#N/A	#N/A
29	Cre01.g053800.v5.5	#N/A	0.020600292	#N/A	#N/A
30	Cre17.g708750.v5.5	#N/A	0.02033964	#N/A	#N/A
31	Cre03.g204250.v5.5	SAH1	0.020274814	S-Adenosyl	S-Adenosyl homocysteine hydrolase, found

				homocysteine hydrolase	in the flagellar proteome [PMID: 15998802]
32	Cre02.g101700.v5.5	#N/A	0.020228653	#N/A	#N/A
33	Cre08.g358527.v5.5	#N/A	0.020208924	#N/A	#N/A
34	**Cre17.g726500.v5.5**	**ORC4**	**0.020197846**	**Origin recognition complex subunit 4**	**Homologous to eukaryotic ORC4, DNA replication control protein**
35	Cre02.g144750.v5.5	PTB4	0.020169258	Sodium/phosphate symporter	Putative phosphate transporter, similar to yeast Pho89 and Neurospora PHO4 probable Na+/Pi symporters; present in a cluster of four highly related PTB family members on scaffold 24, including PTB9, PTB12 and PTB5.
36	Cre04.g215050.v5.5	#N/A	0.020159521	Beta-carotene hydroxylase, putative chloroplast precursor	Beta-carotene hydroxylase (carotene beta-hydroxylase), putative chloroplast precursor (CBH) [PMID:15849308]. Previously annotated as CHYB
37	Cre12.g523900.v5.5	#N/A	0.020138105	#N/A	#N/A
38	Cre14.g618250.v5.5	#N/A	0.020034023	#N/A	#N/A
39	**Cre06.g295700.v5.5**	**MCM3**	**0.020021141**	**Minichromosome maintenance protein**	**Homologous to MCM3 DNA replication protein**
40	**Cre06.g259150.v5.5**	**EFG8**	**0.020004031**	**Mitochondrial translation factor Tu**	**Mitochondrial translation factor Tu**
41	Cre12.g495050.v5.5	#N/A	0.019864311	#N/A	#N/A
42	Cre11.g482700.v5.5	#N/A	0.019848099	#N/A	#N/A
43	Cre03.g205750.v5.5	#N/A	0.01976333	#N/A	#N/A
44	Cre12.g561601.v5.5	#N/A	0.019545801	#N/A	#N/A
45	Cre12.g559800.v5.5	#N/A	0.019527678	#N/A	#N/A
46	Cre16.g688190.v5.5	#N/A	0.019487611	#N/A	#N/A
47	Cre50.g761397.v5.5	#N/A	0.019188946	#N/A	#N/A
48	Cre07.g339250.v5.5	#N/A	0.019184446	#N/A	#N/A
49	Cre10.g441250.v5.5	#N/A	0.019123344	#N/A	#N/A
50	Cre08.g377100.v5.5	ADK6	0.018982981	Adenylate kinase	Putative adenylate kinase
51	Cre09.g407600.v5.5	#N/A	0.018890734	#N/A	#N/A
52	Cre12.g523300.v5.5	GTR22	0.018851218	Glycosyltransferase	Similar to ribophorin I, an essential subunit of oligosaccharyltransferase (OST), a glycosyl transferase which is also known as dolichyl-diphosphooligosaccharide--protein glycosyltransferase [EC 2.4.1.119]; catalyses the transfer of an oligosaccharide fro
53	Cre19.g751047.v5.5	#N/A	0.018683887	#N/A	#N/A
54	**Cre14.g622850.v5.5**	**#N/A**	**0.018644306**	**DNA recombination protein**	**Homologous to RecA and eukaryotic Rad51. Previously annotated as RAD51; homolog of human Rad51A protein [PMID: 17992532]**
55	Cre06.g261300.v5.5	#N/A	0.018594321	#N/A	#N/A
56	Cre14.g611250.v5.5	#N/A	0.018535811	#N/A	#N/A
57	Cre17.g702900.v5.5	#N/A	0.018509732	#N/A	#N/A
58	Cre01.g010296.v5.5	#N/A	0.018406145	#N/A	#N/A
59	Cre12.g551050.v5.5	#N/A	0.01839862	#N/A	#N/A
60	Cre06.g251750.v5.5	#N/A	0.018392712	#N/A	#N/A
61	Cre17.g719900.v5.5	PWD1	0.018385019	Phosphoglucan water dikinase	Catalytic Isoform 3. COG0574, PpsA, Phosphoenolpyruvate synthase/pyruvate phosphate dikinase [Carbohydrate transport and metabolism]. pfam00686, CBM_20, Starch binding domain.
62	Cre02.g092400.v5.5	#N/A	0.018364532	#N/A	#N/A
63	Cre13.g562475.v5.5	#N/A	0.018326516	#N/A	#N/A
64	Cre05.g248150.v5.5	#N/A	0.018320982	#N/A	#N/A
65	Cre01.g032650.v5.5	TAL1	0.018245273	Transaldolase	Putative Transaldolase (EC 2.2.1.2); possible cytosolic isoform, although weak predicted organelle targeting by Target-P; catalyzes the reversible transfer of a three-carbon ketol unit from sedoheptulose 7-phosphate to glyceraldehyde 3-phosphate to form e
66	Cre08.g365900.v5.5	LHCSR1	0.018201948	Stress-related chlorophyll a/b binding protein 1	LI818r-1, stress-related chlorophyll a/b binding protein LhcSR1; Low-CO2 and high-light inducible chlorophyll a/b binding protein, possibly related to photoprotection; regulated by CCM1 [PMID: 15235119]. Acclimation to changing CO2 concentrations and ligh
67	Cre16.g684700.v5.5	#N/A	0.01819795	#N/A	#N/A

68	Cre06.g261450.v5.5	#N/A	0.018152264	#N/A	#N/A
69	Cre17.g731900.v5.5	#N/A	0.018127787	#N/A	#N/A
70	Cre01.g024200.v5.5	#N/A	0.01805345	#N/A	#N/A
71	Cre17.g711150.v5.5	#N/A	0.018030364	#N/A	#N/A
72	Cre06.g278224.v5.5	MRPL16	0.018025769	Putative mitochondrial ribosomal protein L16, imported to mitochondria	Putative mitochondrial ribosomal protein L16
73	Cre12.g538150.v5.5	#N/A	0.017834477	#N/A	#N/A
74	Cre12.g498150.v5.5	#N/A	0.017806527	#N/A	#N/A
75	Cre09.g387875.v5.5	IPY3	0.017687494	Soluble inorganic pyrophosphatase	Soluble inorganic pyrophosphatase; corresponds to CAC42763.1 (ipaU)
76	Cre14.g630859.v5.5	#N/A	0.017661247	#N/A	#N/A
77	Cre12.g550250.v5.5	#N/A	0.017656673	#N/A	#N/A
78	Cre13.g566626.v5.5	#N/A	0.017618414	#N/A	#N/A
79	Cre11.g467641.v5.5	#N/A	0.017518637	#N/A	#N/A
80	Cre01.g026450.v5.5	#N/A	0.017508487	Serine/arginine-rich pre-mRNA splicing factor	Serine/arginine rich (SR) protein involved in mRNA splicing; pre-mRNA splicing factor ASF/SF2; contains two RNA recognition motifs (RRM) Alternative splicing factor ASF/SF2 (RRM superfamily). Previously annotated as SRP35 but this conflicts with signal r
81	Cre16.g678150.v5.5	#N/A	0.017496889	#N/A	#N/A
82	**Cre17.g746347.v5.5**	**UNG1**	**0.017494655**	**Uracil DNA glycosylase**	**Uracil DNA-glycosylase. Prevents mutagenesis by eliminating uracil from DNA by cleaving the N-glycosylic bond and triggering base-excision repair. Uracil bases occur from cytosine deamination or misincorporation of dUMP residues.**
83	**Cre03.g172550.v5.5**	**PRM1**	**0.017491787**	**Protein-/Histone-arginine N-methyltransferase**	**Protein-/Histone-arginine N-methyltransferase; protein arginine N-methyltransferase (ChromDB PRMT3401). Previously annotated as PRMT1. ChromoDB PRMT3401. Ribosomal L11 methyltransferase-like**
84	Cre02.g102450.v5.5	#N/A	0.017304164	#N/A	#N/A
85	Cre06.g249050.v5.5	#N/A	0.017285202	#N/A	#N/A
86	Cre05.g244400.v5.5	#N/A	0.017278621	#N/A	#N/A
87	Cre14.g623176.v5.5	#N/A	0.017270673	#N/A	#N/A
88	Cre01.g019250.v5.5	#N/A	0.017146149	Putative dTDP-glucose 4-6-dehydratase	Putative dTDP-glucose 4-6-dehydratase, an NAD dependent epimerase/dehydratase may be a GDP-mannose 3,5-epimerase. Homolog to Arabidopsis GDP-mannose-3',5'-epimerase (GME) (AT5G28840; PMID: 16366586). Previously annotated as GNE1
89	**Cre12.g558100.v5.5**	**PRM2**	**0.017078951**	**Protein-/Histone-arginine N-methyltransferase**	**Protein-/Histone-arginine N-methyltransferase, ChromDB PRMT3402. Ribosomal L11 methyltransferase-like**
90	Cre06.g298600.v5.5	#N/A	0.017009779	#N/A	#N/A
91	Cre12.g485150.v5.5	GAP1	0.016995427	Glyceraldehyde 3-phosphate dehydrogenase	Glyceraldehyde 3-phosphate dehydrogenase (EC 1.2.1.12), possible cytosolic form; carboxy 90% of deduced AA sequence identical to cytosolic G3PC_CHLRE [Genbank P49644; PMID: 8114942], but the N-terminal AA's do not match those of G3PC_CHLRE. Two splice va
92	**Cre12.g515850.v5.5**	**PCN1**	**0.016992389**	**Proliferating cell nuclear antigen homolog**	**Proliferating Cell Nuclear Antigen (PCNA) homolog**
93	Cre03.g190800.v5.5	#N/A	0.016990526	#N/A	#N/A
94	Cre17.g705550.v5.5	#N/A	0.016962485	#N/A	#N/A
95	Cre09.g403071.v5.5	#N/A	0.016948114	#N/A	#N/A
96	Cre16.g679350.v5.5	#N/A	0.016935673	#N/A	#N/A
97	Cre07.g318276.v5.5	#N/A	0.016915499	#N/A	#N/A
98	Cre16.g684850.v5.5	#N/A	0.016910753	#N/A	#N/A
99	Cre06.g282850.v5.5	#N/A	0.016907286	#N/A	#N/A
100	Cre15.g635400.v5.5	ZYS3	0.016902683	Zygote-specific protein	Zygote-specific protein 3; partial sequence contains ankyrin repeats. Previously annotated as ZYS1

Appendix 5. List of the 100 most positive expressed genes (*iclC* phenotype) of the second constraint when 8 dark samples are analysed separately. *C. reinhardtii* genome version v5.5. #N/A means the function of the gene is unknown.

Dark samples – constraint 2 – 100 most negative genes- *icl* phenotype					
	Identifier	Name	G_{i2}value	Definition	Description
1	**Cre04.g223100.v5.5**	**CAH1**	**-0.093491136**	**Carbonic anhydrase**	**Carbonate dehydratase 1 in Chlamydomonas reinhardtii, also known as CA1; gi:115447; Carbonic anhydrase 1 precursor (Carbonate dehydratase 1) (CA1); alpha type - Carbonic anhydrase 1 localized in the periplasmic space and low-CO2 inducible gene regulated b**
2	**Cre09.g399552.v5.5**	**LCR1**	**-0.080389229**	**Low-CO2 response regulator, Myb-like transcription factor**	**Myb-DNA binding transcriptional factor induced under low CO2 stress condition [PMID: 15155888]; regulates CO2-responsive genes, Cah1, Lci1, and Lci6; regulated by CCM1 [PMID: 15235119]**
3	Cre16.g674151.v5.5	#N/A	-0.074475703	#N/A	#N/A
4	Cre17.g723400.v5.5	#N/A	-0.072628461	#N/A	#N/A
5	Cre26.g756747.v5.5	#N/A	-0.066579144	#N/A	#N/A
6	Cre12.g554929.v5.5	#N/A	-0.064668763	#N/A	#N/A
7	*Cre17.g723350.v5.5*	*SUL2*	*-0.064191154*	*Sulfate anion transporter*	*Similar to Arabidopsis Sultr4;1. Previously annotated as SULTR2*
8	Cre16.g681351.v5.5	#N/A	-0.062693595	#N/A	#N/A
9	Cre06.g298802.v5.5	#N/A	-0.062194566	#N/A	#N/A
10	Cre07.g342551.v5.5	#N/A	-0.061800919	#N/A	#N/A
11	Cre09.g392208.v5.5	#N/A	-0.061143349	#N/A	#N/A
12	*Cre02.g144700.v5.5*	*PTB5*	*-0.060907616*	*Sodium/phosphate symporter*	*Putative phosphate transporter, similar to yeast Pho89 and Neurospora PHO4 probable Na+/Pi symporters; present in a cluster of four highly related PTB family members on scaffold 24, including PTB9, PTB12 and PTB4.*
13	Cre13.g579650.v5.5	#N/A	-0.058428136	#N/A	#N/A
14	Cre07.g328800.v5.5	NSG13	-0.05767398	Protein expressed during nitrogen-starved gametogenesis	Predicted to contain 3 transmembrane helices and 4 Armadillo repeats; expressed exclusively during gametogenesis, at a late stage (PMID: 15459796). Identified by Abe et al. 2004 [PMID: 15459796] as NSG13, a gene expressed during nitrogen-starved gametoge
15	Cre05.g234661.v5.5	BCS1	-0.057501182	Ubiquinol:cytochrome c oxidoreductase biogenesis factor	Related to ubiquinol:cytochrome c oxidoreductase (Complex III) BCS1 biogenesis factor; similar to yeast Bcs1p mitochondrial protein required for expression of functional Rieske iron-sulfur protein; region related to AAA family ATPase [PMID: 15710684]
16	*Cre12.g502600.v5.5*	*SLT1*	*-0.057280422*	*Sodium/sulfate co-transporter*	*SAC1-like transporter 1, putative sodium/sulfate co-transporter, transcript is up-regulated during sulfur-deprivation; related to the SAC1 protein which regulates sulfur-deficiency responses [PMID: 16307308]*
17	Cre10.g447800.v5.5	#N/A	-0.057055325	#N/A	#N/A
18	Cre17.g725750.v5.5	#N/A	-0.05560777	#N/A	#N/A
19	Cre05.g236650.v5.5	#N/A	-0.05502286	#N/A	#N/A
20	Cre03.g163950.v5.5	CDO2	-0.05495774	Cysteine dioxygenase	Cysteine dioxygenase
21	Cre03.g151650.v5.5	#N/A	-0.054417992	#N/A	#N/A
22	**Cre09.g394473.v5.5**	**LCI9**	**-0.054322703**	**Low-CO2-inducible protein**	**Low-CO2 inducible protein, containing starch-binding domain of CBM_20 (pfam00686) found in glycoside hydrolases. Because Chlamydomonas cells develop pyrenoids under low-CO2 conditions, which are surrounded by a typical starch sheath, Lci9 may function in**
23	Cre03.g155150.v5.5	#N/A	-0.053154071	#N/A	#N/A
24	Cre06.g293100.v5.5	#N/A	-0.05294192	#N/A	#N/A
24	Cre12.g551700.v5.5	#N/A	-0.052816654	#N/A	#N/A
26	Cre10.g431900.v5.5	#N/A	-0.052506242	#N/A	#N/A
27	Cre06.g301750.v5.5	#N/A	-0.052253932	#N/A	#N/A
28	**Cre16.g663450.v5.5**	**#N/A**	**-0.051479239**	**Low-CO2-inducible membrane protein**	**Low-CO2 inducible gene revealed by cDNA analyses. At least four other paralogous genes are conserved in the Chlamydomonas genome. Predicted amino acid sequence is homologous to Arabidopsis hypothetical protein T20K12.220. SOSUI prediction: 1 transmembrane**
29	Cre10.g429000.v5.5	#N/A	-0.051342956	#N/A	#N/A
30	Cre13.g570801.v5.5	#N/A	-0.050076615	#N/A	#N/A
31	Cre11.g468050.v5.5	#N/A	-0.050017061	#N/A	#N/A

32	Cre12.g505100.v5.5	#N/A	-0.049886782	#N/A	#N/A
33	Cre09.g403550.v5.5	#N/A	-0.04968069	#N/A	#N/A
34	Cre07.g342552.v5.5	#N/A	-0.049625724	#N/A	#N/A
35	Cre12.g505050.v5.5	#N/A	-0.049598443	#N/A	#N/A
36	Cre07.g333350.v5.5	#N/A	-0.049417794	#N/A	#N/A
37	Cre12.g518800.v5.5	#N/A	-0.049047847	#N/A	#N/A
38	Cre10.g428966.v5.5	#N/A	-0.048729978	#N/A	#N/A
39	Cre06.g287350.v5.5	#N/A	-0.048595837	#N/A	#N/A
40	Cre16.g662600.v5.5	#N/A	-0.04851098	#N/A	#N/A
41	Cre02.g106450.v5.5	#N/A	-0.048162391	#N/A	#N/A
42	Cre05.g239450.v5.5	#N/A	-0.047876842	#N/A	#N/A
43	Cre01.g029000.v5.5	#N/A	-0.047597183	Ubiquinone/menaquinone biosynthesis methyltransferase	UbiE/COQ5 methyltransferase family; possibly functioning in chloroplast. Previously annotated as COQ5D
44	Cre16.g681400.v5.5	#N/A	-0.047560059	#N/A	#N/A
45	Cre12.g485600.v5.5	#N/A	-0.047459103	#N/A	#N/A
46	Cre14.g617400.v5.5	HSP22F	-0.047257528	Heat shock protein 22F	Small heat shock protein containing an alpha-crystallin domain; N-terminus by TargetP weakly predicted to be chloroplast transit peptide; HSP22F and HSP22E are arranged head-to-head and share 98% identical nucleotide sequence in coding region (96% identic
47	Cre13.g585000.v5.5	#N/A	-0.04718222	#N/A	#N/A
48	Cre12.g495952.v5.5	#N/A	-0.04692837	#N/A	#N/A
49	Cre02.g107000.v5.5	#N/A	-0.046881033	#N/A	#N/A
50	Cre01.g004750.v5.5	#N/A	-0.046680195	#N/A	#N/A
51	Cre06.g278107.v5.5	#N/A	-0.04665122	#N/A	#N/A
52	Cre05.g238311.v5.5	#N/A	-0.0463229	#N/A	#N/A
53	Cre03.g195950.v5.5	#N/A	-0.046027939	#N/A	#N/A
54	Cre08.g367500.v5.5	LHCSR2	-0.045988978	Stress-related chlorophyll a/b binding protein 2	Stress-related chlorophyll a/b binding protein, similar to LI8I8r-1/LHCSR1. Peers et al. (2009) [PMID: 19940928] call this gene LHCSR3.1; this apparently codes for the same protein as LHCSR3.2
55	Cre13.g569450.v5.5	#N/A	-0.045930887	#N/A	#N/A
56	Cre16.g686203.v5.5	#N/A	-0.045486293	#N/A	#N/A
57	Cre07.g352550.v5.5	RDP3	-0.044974058	Putative rhodanese domain phosphatase	Putative rhodanese domain phosphatase similar to CDC25; M-phase inducer phosphatase. Putative function in G2/M
58	Cre13.g569500.v5.5	#N/A	-0.044376068	#N/A	#N/A
59	Cre13.g569400.v5.5	#N/A	-0.04425879	#N/A	#N/A
60	Cre02.g144734.v5.5	#N/A	-0.044105757	#N/A	#N/A
61	**Cre03.g162800.v5.5**	**LCI1**	**-0.043951652**	**Low-CO2-inducible membrane protein**	**Low-CO2-inducible gene encoding 21 kDa transmembrane protein; Burow et al. 1996 [PMID: 8756610]; SOSUI prediction: signal peptide in N-terminal end and 3 transmembrane helices. Regulated by CO2 through MYB-transcription factor LCR1 [PMID: 15155888] and CC**
62	Cre13.g605500.v5.5	#N/A	-0.043521157	#N/A	#N/A
63	Cre14.g617450.v5.5	HSP22E	-0.0434184	Heat shock protein 22E	Small heat shock protein containing an alpha-crystallin domain; N-terminus by TargetP weakly predicted to be chloroplast transit peptide; HSP22E and HSP22F are arranged head-to-head and share 98% identical nucleotide sequence in coding region (96% identic
64	Cre02.g093750.v5.5	NRX2	-0.043376863	Nucleoredoxin 2	Nucleoredoxin, nuclear thioredoxin-like protein, contains two thioredoxin motifs with a WCPPC active site in each motif.
65	Cre11.g477350.v5.5	#N/A	-0.043269827	#N/A	#N/A
66	Cre03.g201417.v5.5	#N/A	-0.043102993	#N/A	#N/A
67	Cre06.g310950.v5.5	#N/A	-0.043036169	#N/A	#N/A
68	Cre15.g642865.v5.5	#N/A	-0.042849913	#N/A	#N/A
69	Cre07.g343050.v5.5	#N/A	-0.042222769	#N/A	#N/A
70	Cre06.g311050.v5.5	#N/A	-0.041784658	#N/A	#N/A
71	Cre13.g583450.v5.5	#N/A	-0.041742481	#N/A	#N/A
72	Cre04.g214750.v5.5	#N/A	-0.041368176	Endosomal R-SNARE protein, VAMP-like family (R.III)	Endosomal R-SNARE protein, VAMP-like family (R.III). Expressed Protein. Contains R-SNARE domain distantly similar to VAMP4, not found in land plants. Part of 2 member gene family on same scaffold separated by one gene (<-VMPL1, IFM1, <-VMPL2). Class R.I

73	Cre07.g355500.v5.5	#N/A	-0.041061077	#N/A	#N/A
74	*Cre02.g095076.v5.5*	*MFT10*	*-0.04021291*	*Major facilitator superfamily transporter, involved in circadian rhythm control*	*Major facilitator superfamily transporter. Matsuo et al.'s (2008) roc39 (rhythm of chloroplast) circadian rhythm mutant locus maps here*
75	Cre11.g467531.v5.5	FAP15	-0.039884392	Flagellar Associated Protein	Flagellar Associated Protein, found in the flagellar proteome [PMID: 15998802]
76	*Cre06.g309000.v5.5*	*NAR1.2*	*-0.039835481*	*Anion transporter*	*Nitrite transporter-homologue, possibly chloroplastic [PMID: 11912227, AY612639]; significant similarities with two anion transporters, the chloroplastic nitrite transporter NAR1 and formate transporters in bacteria; also known as LCIA. nduced by CO2-limi*
77	*Cre06.g298750.v5.5*	*AOT4*	*-0.039685863*	*Amino acid transporter*	*Belongs to AAAP family of amino acid/auxin permeases; most similar to plant and fungal relatives*
78	Cre08.g366700.v5.5	#N/A	-0.039448194	#N/A	#N/A
79	**Cre06.g281600.v5.5**	**LCI23**	**-0.03942349**	**Low-CO2-inducible protein, septin-like**	**Low-CO2 inducible gene revealed by cDNA array analyses; regulated by CCM1 [PMID: 15235119]. Acclimation to changing CO2 concentrations and light intensities was studied by Yamano et al. 2008 [PMID: 18322145], appears in their table as Lci22 rather than LC**
80	Cre13.g569150.v5.5	#N/A	-0.039397034	#N/A	#N/A
81	Cre02.g077750.v5.5	FAP211	-0.03870336	Flagellar Associated Protein	Flagellar Associated Protein, found in the flagellar proteome [PMID: 15998802]
82	Cre06.g278098.v5.5	MCC1	-0.038651834	Methylcrotonoyl-CoA carboxylase alpha subunit	Methylcrotonoyl-CoA carboxylase subunit alpha, mitochondrial precursor; 3-methylcrotonyl-CoA carboxylase; biotin-dependent carboxylase. Previously annotated as MCCA. Contains a very short LSU-rRNA motif on the opposite strand (4456221-4456270), in intro
83	Cre09.g393953.v5.5	#N/A	-0.038578841	#N/A	#N/A
84	Cre08.g367400.v5.5	LHCSR3	-0.038514599	Stress-related chlorophyll a/b binding protein 3	Stress-related chlorophyll a/b binding protein, similar to LI8I8r-1/LHCSR1. Peers et al. (2009) [PMID: 19940928] call this gene LHCSR3.2; this apparently codes for the same protein as LHCSR3.1
85	Cre06.g259550.v5.5	#N/A	-0.038245634	#N/A	#N/A
86	Cre12.g504900.v5.5	#N/A	-0.038184172	#N/A	#N/A
87	Cre03.g176833.v5.5	#N/A	-0.038122396	#N/A	#N/A
88	Cre02.g095080.v5.5	#N/A	-0.037415358	#N/A	#N/A
89	Cre16.g685000.v5.5	#N/A	-0.037367527	#N/A	#N/A
90	Cre01.g053288.v5.5	#N/A	-0.037221385	#N/A	#N/A
91	Cre11.g467693.v5.5	#N/A	-0.037187724	#N/A	#N/A
92	Cre12.g554400.v5.5	#N/A	-0.037046915	#N/A	#N/A
93	Cre16.g694500.v5.5	#N/A	-0.036920634	DnaJ-like protein	DnaJ-like protein
94	Cre09.g396809.v5.5	#N/A	-0.036814547	#N/A	#N/A
95	Cre04.g215900.v5.5	#N/A	-0.03643577	#N/A	#N/A
96	Cre12.g516700.v5.5	#N/A	-0.036166955	#N/A	#N/A
97	Cre04.g217925.v5.5	#N/A	-0.036123543	#N/A	#N/A
98	Cre02.g093800.v5.5	NRX3	-0.036063993	Nucleoredoxin	Related to animal thioredoxin, nucleoredoxin
99	Cre03.g144144.v5.5	#N/A	-0.03598311	#N/A	#N/A
100	Cre17.g734725.v5.5	#N/A	-0.035441674	#N/A	#N/A

Appendix 6. List of the 100 most negative expressed genes (*icl* phenotype) of the second constraint when dark samples are analysed separately. *C. reinhardtii* genome version v5.5. #N/A means the function of the gene is unknown.