New features on the environmental regulation of metabolism revealed by modeling the cellular proteomic adaptations induced by light, carbon and inorganic nitrogen in *Chlamydomonas reinhardtii*

Stéphanie GÉRIN[1], Pierre LEPRINCE[2], Francis E SLUSE[1], Fabrice FRANCK[1*], Grégory MATHY [3*]

[1]Laboratory of Bioenergetics, Department of Life Sciences, Faculty of Sciences, University of Liege, Liege, Belgium

[2]Laboratory of Nervous System Disorders and Therapy, GIGA-Neurosciences, Faculty of Medicine, University of Liege, Liege, Belgium

[3]Upstream Process Sciences, UCB Pharma, Braine l'Alleud, Belgium

***Correspondence:**
Fabrice FRANCK
F.Franck@ulg.ac.be
Grégory MATHY
Gregory.Mathy@ucb.com

Number of words: 11,966
Number of figures: 7
Number of tables: 5

Keywords

2D-DIGE, Design of experiments, Hierarchical clustering, Multiple linear regression, Bioenergetics, Metabolic network, Biological system, Environment

Abstract

Microalgae are currently emerging to be very promising organisms for the production of biofuels and high-added value compounds. Understanding the influence of environmental alterations on their metabolism is a crucial issue. Light, carbon and nitrogen availability have been reported to induce important metabolic adaptations. So far, the influence of these variables has essentially been studied while varying only one or two environmental factors at the same time. The goal of the present work was to model the cellular proteomic adaptations of the green microalga *Chlamydomonas reinhardtii* upon the simultaneous changes of light intensity, carbon concentrations (CO_2 and acetate) and inorganic nitrogen concentrations (nitrate and ammonium) in the culture medium. Statistical design of experiments (DOE) enabled to define 32 culture conditions to be tested experimentally. Relative protein abundance was quantified by two dimensional differential in-gel electrophoresis (2D-DIGE). Additional assays for respiration, photosynthesis, and lipid and pigment concentrations were also carried out. A hierarchical clustering survey enabled to partition biological variables (proteins + assays) into eight co-regulated clusters. In most cases, the biological variables partitioned in the same cluster had already been reported to participate to common biological functions (acetate assimilation, bioenergetic processes, light harvesting, Calvin cycle and protein metabolism). The environmental regulation within each cluster was further characterized by a series of multivariate methods including principal component analysis and multiple linear regressions. This metadata analysis enabled to highlight the existence of a clear regulatory pattern for every cluster and to mathematically simulate the effects of light, carbon and nitrogen. The influence of these environmental variables on cellular metabolism is described in details and thoroughly discussed. This work provides an overview of the metabolic adaptations contributing to maintain cellular homeostasis upon extensive environmental changes. Some of the results presented here could be used as starting points for more specific fundamental or applied investigations.

1. Background

Freshwater green microalgae are known to undergo global metabolic reorganizations to adapt to changing environmental conditions. This enables microalgae to maintain their cellular homeostasis despite the onset of very dynamic modifications of physico-chemical parameters such as temperature, nutrient availability or gas partial pressures (Falkowski and Raven, 2013). *Chlamydomonas reinhardtii* is a model organism which is commonly used to study photosynthetic processes. This green microalga exhibits a much faster growth rate than higher plants, is of easier maintenance and can be cultured under very diverse experimental conditions (Harris, 2001). *C. reinhardtii* is able to grow either in the light (photoautotrophy in the presence of CO_2; mixotrophy in the presence of CO_2 + organic carbon) or in the dark when an oxidizable carbon source is available in the medium (Spalding, 2009;Perez-Garcia et al., 2011). Moreover it can assimilate different chemical forms of nitrogen, either inorganic (nitrate, nitrite, ammonium) or organic (urea, amino acids, purine nucleotides) (Fernandez et al., 2004). In 2007, the sequencing of the *C. reinhardtii* genome opened the gate to further characterization by a growing panel of molecular techniques such as targeted mutagenesis, transcriptomics and proteomics (Merchant et al., 2007).

Over the last decades, the influence of environmental changes on biological functions has been extensively studied in photosynthetic organisms. Light, carbon and nitrogen have been reported to induce dramatic metabolic adaptations as a way to maintain a proper bioenergetic balance. These adaptations can occur at very different levels such as genetic expression, protein abundance, enzymatic activity or cellular structure (Tobin and Silverthorne, 1985;Spalding et al., 2002;Fernandez et al., 2004). To date, most studies have focused on the individual effects of light, carbon and nitrogen (light + carbon or nitrogen + carbon in a few cases). However, from available data, it is clear that their signaling and assimilatory pathways are connected through a complex metabolic network (Turpin, 1991;Huppe and Turpin, 1994;Singh et al., 2008). Understanding how photosynthetic organisms adapt to global environmental modifications could therefore be of prime interest. This is all the more true that green microalgae are currently emerging as very promising sources for the production of biofuels and high-added value compounds (Work et al., 2012). In this context, the bioenergetic adaptations of *C. reinhardtii* cells upon simultaneous changes related to light, carbon and inorganic nitrogen have recently been modeled by our group (Gérin et al., 2014). This work was carried out through a statistical approach coupling design of experiments (DOE) to multiple linear regression analyses. It enabled to build empirical models simulating mathematically the influence of each environmental variable and highlighting significant interactions between them in some cases.

Comparative proteomics is a suitable tool to characterize the metabolic adaptations induced by diverse endogenous or exogenous perturbations such as genetic modifications, pathologies, heat shocks or nutrient changes. In this field, proteomics is often preferred to transcriptomics since the correlation between mRNA abundance and protein expression is relatively weak, as reviewed in (Greenbaum et al., 2003) for yeast. 2D-DIGE (two dimensional differential in-gel electrophoresis) is a comparative proteomic technique requiring the pre-electrophoretic labeling of protein samples with three spectrally-distinct fluorescent dyes. Its capacity for multiplexing enables to introduce an internal standard in each gel electrophoresis to normalize protein abundance. Together with the relatively wide dynamic range of 2D-DIGE, this makes it possible to quantify very accurate changes in protein abundance (Marouga et al., 2005).

To date, an overview of the influence of cumulative environmental changes on photosynthetic metabolism is lacking. Some pathways have nevertheless been reported (mostly through univariate studies) to be regulated by two or several factors: see for example the well-known regulation of photosynthetic antennae size and pigment content by light and acetate, and the respective effects of acetate and inorganic nitrogen on the TCA cycle (Neale and Melis, 1986;Falkowski and LaRoche, 1991;Turpin, 1991;Huppe and Turpin, 1994;Teramoto et al., 2002;Durnford et al., 2003;Nield et al., 2004;Boyle and Morgan, 2009;Gérin et al., 2010;Gérin et al., 2014). In the present work, we aimed to build statistical models describing the global metabolic adaptations of *C. reinhardtii* cells upon simultaneous changes of several environmental variables: light intensity, carbon concentration (acetate and CO_2) and inorganic nitrogen concentration (nitrate and ammonium). Such an empirical approach appeared to us as a crucial pre-requisite before attempting to build mechanistic models in subsequent studies. For this purpose, 2D-DIGE was coupled to DOE and multivariate data analyses in order to characterize the environmental regulation of protein abundance at the cellular level. Additional assays for respiration, photosynthesis and cellular contents of some lipids and pigments were also carried out to this end. Hierarchical clustering was first performed to partition biological variables (proteins and assays) into discrete co-regulated clusters. The pattern of environmental regulation was then characterized within each cluster through a panel of multivariate statistical methods, including principal component analysis and multiple linear regressions. Overall, the data reported here provide an overview of the metabolic adaptations set up in response to global environmental changes related to light, carbon and inorganic nitrogen in *C. reinhardtii*.

2. Material and Methods

2.1 Cell cultures

A *cw15* mt^+ wall-less strain of *C. reinhardtii* (Hyams and Davies, 1972) was used in this study (*Chlamydomonas* Resource Center ID: CC-400). Algal cells were cultivated in lab-scale tubular photobioreactors (Multi-Cultivators MC 1000, Photon System Instruments) as described in (Gérin et al., 2014) (same media, conditions and procedures). Culture media invariably contained $MgSO_4$ 1.4 mM, $CaCl_2$ 450 µM, K_2HPO_4 5.4 mM, KH_2PO_4 4.6 mM, Tris-HCl 20 mM pH 7.2, added with oligo-elements (composition described in (Gérin et al., 2014)). When applicable according to the DOE, acetic acid (0 to 1 $g.L^{-1}$), $NaNO_3$ (0 to 20 mM) and NH_4Cl (0 to 15 mM) were also added. CO_2 was bubbled in the cultures at either 0.035% (ambient air) or 1.5% (mix of ambient air + pure CO_2). Light intensity was tuned from 0 to 200 $\mu mol_{photons}.m^{-2}.s^{-1}$ with the Multi-Cultivator interface. Algae were harvested by centrifuging at 3,000 *g* for 5 min, washed in one volume of ice-cold saline buffer (NaCl 150 mM, Tris-HCl 50 mM, pH 7.2), centrifuged again in the same conditions and stored as pellets at -80°C for analyses.

2.2 2D-DIGE

2.2.1 Protein extraction and purification

Algal pellets were resuspended in an ice-cold extraction buffer (NaCl 150 mM, Triton X-100 0.1% (v/v), EDTA 1 mM, DL-dithiothreitol (DTT) 25 mM, complete EDTA-free protease inhibitor cocktail tablets (Roche), Tris-HCl 50 mM pH 7.8) added with

polyvinylpolypyrrolidone (PVPP, insoluble in water) 2.5% (w/v) to complex polyphenols. Proteins were extracted by sonicating at 6 Amp for 30 s on ice (Sonifier Cell Disruptor B-12, Branson), vortexing for 30 s at 4°C and repeating the procedure twice more. Protein extracts were centrifuged at 3,000 *g* for 3 min at 4°C to spin down PVPP. The supernatant was centrifuged again at 10,000 *g* for 3 min to spin down cellular debris, and was then filtered with a 0.22 µm cellulose acetate-membrane syringe filter. Proteins were further purified according to the phenol phase separation procedure described by Carpentier and co-workers (Carpentier et al., 2005), and were finally solubilized in an appropriate volume of a DIGE labeling buffer (urea 7 M, thiourea 2 M, ASB-14 2% (w/v), EDTA 0.5 mM, DTT 10 mM, Tris-HCl 50 mM pH 8.5) so as to reach a concentration comprised between 5 and 10 $mg.mL^{-1}$.

2.2.2 Protein labeling

Protein samples were labeled with Refraction-2D G-Dyes from NH DyeAgnostics (May et al., 2012) and allocated to 16 different 2D-electrophoreses as detailed in Additional file 1. Each electrophoresis comprised two algal culture samples labeled with G-Dye200 and G-Dye300, and an internal standard (equal amount of all available samples) labeled with G-Dye100. In each case, 25 µg of proteins were labeled with 0.2 nmol of G-Dye for 30 min at 25°C in the dark. Labeling reactions were stopped by adding 1 µL of Stop Solution (DyeAgnostics) and incubating samples for 10 min in the same conditions. For preparative electrophoreses, a 500 µg pool of all samples in an equal amount was constituted, out of which 25 µg were labeled with G-Dye100 before being re-incorporated among the remaining 475 µg.

2.2.3 2D-electrophoreses and image acquisition

Isoelectrofocusing (IEF) was carried out as previously reported (Mathy et al., 2010) by using a 3-11 non-linear pH range, except that the rehydration buffer was added with CHAPS 3% (w/v). IPG strips (GE Healthcare) were then rinsed with milliQ water before being reduced, alkylated and loaded on the top of polyacrylamide gels for SDS-PAGE separation as also described in this previous publication (Mathy et al., 2010), with the difference that 10% polyacrylamide gels (37.5:1 acrylamide-to-bisacrylamide ratio) were rather used in the present study. Images of G-Dyes within 2D-gels were acquired with a Typhoon 9400 scanner (GE Healthcare) by using the specific excitation and emission wavelength of each dye.

2.2.4 Image analysis

Images were analyzed with the DeCyder 7.0 software from GE Healthcare. Spot detection was performed in the Differential In-Gel Analysis (DIA) module with an exclusion filter restricting detection to protein spots with a volume superior or equal to 300,000 (for at least one G-Dye over three). The Biological Variation Analysis (BVA) module was then used to perform inter-gel matching of protein spots (Match Table) and to extract the abundance normalized by the internal standard for each spot and each culture condition (Appearance Table). These abundance values are the raw data used in the present statistical analyses.

2.2.5 Spot picking and protein identification by mass spectrometry

Protein spots were picked off preparative gels with an Ettan DALT Spot Picker device (GE Healthcare) and in-gel digested according to Shevchenko and co-workers (Shevchenko et

al., 1996). Peptides were then extracted from gel pieces and prepared for mass spectrometry as previously described (Mathy et al., 2010).

Acquisition of mass spectra was carried out with a MALDI-TOF/TOF mass spectrometer (Ultraflex II, Bruker Daltonics) in PMF mode. The device was piloted by FlexControl 3.0, with real-time analysis of mass spectra by FlexAnalysis 3.0 and database search by BioTools 3.1 in the Mascot server, version 2.2.04. Database search was performed in NCBInr restricted to *Viridiplantae* (1,930,642 sequences) with the Mascot PMF algorithm as search engine and 100 ppm of mass error tolerance. Cysteine carbamidomethylation and methionine oxidation were assessed as fixed and variable peptide modifications, respectively. Protein identification was considered as successful for Mascot scores equal or superior to 75. Protein function(s) and cellular location(s) were searched in the ChlamyCyc database, version 1.0 (May et al., 2009).

2.3 Pigment extraction and analysis

For the determination of chlorophyll *a*, chlorophyll *b* and total carotenoid (TC) contents, the absorbance of pigment extracts in methanol was measured at 470, 652 and 665 nm, and the Lichtenthaler and Wellburn's formulas (Lichtenthaler and Wellburn, 1983) were applied.

Neoxanthin, lutein, violaxanthin and β-carotene concentrations were determined by high pressure liquid chromatography (HPLC) by using pigment extracts in methanol. Pigments were separated in a Nova-Pak silica-based, reverse-phase, 4 µm particle C_{18} column (Waters, product WAT036975, length: 150 mm, inner diameter: 3.9 mm). HPLC experiments were run with a 1 $mL.min^{-1}$ flow rate at 25°C with three working solutions: solution A (methanol 90%, ammonium acetate 100 mM), solution B (acetonitrile 90%) and solution C (ethyl acetate 100%). The following protocol was applied as gradient: 0 min - 100% A; 0.5 min - 100% B; 1.1 min - 90% B + 10% C; 6.1 min - 65% B + 35% C; 11.5 min - 40% B + 60% C; 15.0 min - 100% C; 17.0 min - 100% A; 23.0 min - 100% A. Pigment elution times were determined by using the Mixed phytoplankton pigment standard (PPS-MIX-1) from DHI Lab Products. Chromatograms were analyzed at 430 nm and relative pigment concentrations were assessed in terms of peak areas at this wavelength.

2.4 Fatty acid extraction and analysis

Fatty acids were extracted with chloroform-methanol and transesterified as previously described (Bligh and Dyer, 1959;Browse et al., 1986). Fatty acid concentrations were determined by gas chromatography (GC) with a BPX70 70% cyanopropylpolysilphenylene-siloxane column (SGE Analytical Science, product 054622, length: 2m, inner diameter: 0.25 mm) with helium as carrier gas at 250°C. FAMES elution times and calibration curve were determined by running the Supelco 37 Component FAMES mix standard from Sigma-Aldrich (product CRM47885).

2.5 Triglyceride extraction and analysis

Algal pellets were resuspended in an extraction buffer (NaCl 150 mM, Triton X-100 0.1% (v/v), Tris-HCl 50 mM pH 7.5), sonicated at 3 Amp for 15 s (Sonifier Cell Disruptor B-12, Branson) and thoroughly vortexed for 20 min. Triglyceride concentration was determined by using the enzymatic assay kit of BioVision (product K622-100).

2.6 Protein assay for 2D-DIGE and pigment/lipid normalization

Protein concentrations were determined by using the Reagent Compatible/Detergent Compatible assay kit from BioRad (product 500-0121) which is based on the Lowry-Ciocalteu colorimetric method (Lowry et al., 1951).

2.7 Respiratory and photosynthetic parameters

Bioenergetic data were extracted from our previous modeling publication (Gérin et al., 2014). As described there, these data were obtained by oxymetric measurements and pulse-amplitude-modulated (PAM) fluorimetry. Photosynthetic parameters ($\varphi PSII_{800}$, P_{800} and NPQ_{800}) were measured under a saturating light irradiance of 800 $\mu mol_{photons}.m^{-2}.s^{-1}$.

2.8 Statistical analyses

2.8.1 Design of experiments

Design of experiments (DOE) was carried out with the Custom design platform of the JMP 11 software (SAS) with the following parameters: one dependent variable (goal: none); five environmental variables (changes: easy); single effects, 2nd-degree polynomial effects (for continuous factors) and 2nd-order interactions (estimability: necessary); 7 center points, zero replicate runs, default number of assays; randomize output order. Light intensity, nitrogen concentrations and acetate concentration were considered as continuous variables, whereas CO_2 concentration was considered as an ordinal variable with two modalities. For each continuous environmental variable, the minimal value was set to zero. The maximal values (described in point 2.1) were chosen as follow:

- for nitrogen and acetate concentrations: twice higher than the optimal level - this generates values that enable biomass accumulation but remain below toxicity (Sager and Granick, 1953;Chen and Johns, 1994; 1996;Collos and Harrison, 2014;Gérin et al., 2014)
- for light intensity: insufficient to saturate the photosynthetic apparatus in order to limit photo-oxidative damages during algal cultivation (Sueltemeyer et al., 1986;White and Critchley, 1999)

As described in point 2.1, the two modalities of CO_2 concentration were set at its atmospheric level (0.035%) and at a saturating concentration (1.5%) sufficient to ensure no CO_2 limitation for RubisCO whatever the light intensity (Vance and Spalding, 2005).

2.8.2 Initial screening for biological variables

An initial screening of the biological variables relevantly influenced by one or several environmental variable(s) was carried out with the JMP 11 software (SAS) by a methodology coupling PLSR and MLR as detailed in Table 1. PLSRs were run in the Mulivariate methods platform through the NIPALS algorithm with selection of the Centering and Scaling options. Leave-one-out validation method was chosen, and the default factor search range displayed by the software was not modified. The optimal number of latent factors was determined by using minimal PRESS (prediction error sum of squares) coupled to van der Voet T^2 tests as selection criteria. MLRs were run in the Fit model platform. Screening was performed independently for protein spots and additional assays.

2.8.3 Hierarchical clustering

Hierarchical clustering was performed in the Multivariate methods platform of the JMP 11 software (SAS) by the Ward's minimum variance method (Ward Jr, 1963;SAS, 2013). The options "Standardize data" and "Missing value imputation" were selected. The imputation of missing values was performed as follow: a single covariance matrix was built by the pairwise method on the basis of the whole data set; the nonmissing variables were then used as predictors to impute missing values by a method equivalent to regression prediction (SAS, 2013).

2.8.4 Gene set enrichment analyses

Gene set enrichment analyses were performed in the PANTHER (Protein ANalysis THrough Evolutionary Relationships) database. The GI numbers in NCBI were used as protein IDs. *Chlamydomonas reinhardtii* was selected as organism. The "PANTHER Overrepresentation test (release 20160321)" was used as analysis type. The annotation data set was either "PANTHER Pathways" (PANTHER version 10.0 Released 2015-05-15) or "GO cellular component complete" (GO Ontology database Released 2016-05-20). The *p*-values were extracted with and without Bonferroni correction for multiple testing.

2.8.5 Principal component analysis (PCA) and in-cluster PLSRs

These procedures were both performed in the Multivariate methods platform of the JMP 12 software (SAS) with standardized data (i.e. data scaled to a mean of 0 and centered to a variance of 1 for each biological and environmental variable). PCA was carried out by the pairwise method on the basis of the correlation matrix with all biological variables in the same PCA. In-cluster PLSRs were performed through the NIPALS algorithm with all five environmental variables as factors. For CO_2 concentration, data were first transformed according to a binary code: 0 for the lowest modality (0.035% CO_2) and 1 for the highest one (1.5% CO_2). Leave-one-out validation method was chosen, and the default factor search range displayed by the software was set to five. The optimal number of latent factors was determined by using minimal PRESS (prediction error sum of squares) coupled to van der Voet T^2 tests as selection criteria. All biological variables (proteins + assays) belonging to each cluster were included in the same PLSR analysis.

2.8.6 Multiple correlations

Multiple correlations were assessed in the Multivariate methods platform of the JMP 11 software (SAS). The Pearson's correlation coefficients (R) between biological variables and the corresponding *p*-values were calculated by the pairwise method without missing value imputation. Data were previously centered to a mean of 0 and scaled to a variance of 1 for every biological variable before being analyzed.

2.8.7 Modeling the dependence of biological variables upon environmental variables

Modeling was performed in the Fit model platform of the JMP 11 software (SAS) on the basis of raw data listed in Additional file 2, following the same procedure as previously described (Gérin et al., 2014). Linear effects, quadratic effects and second-order interactions of the environmental variables were assumed.

2.8.7.1 Stepwise regression

Stepwise regression was carried out in forward direction with minimum AICc (corrected Akaike information criterion) as stopping rule (Burnham and Anderson, 2004).

2.8.7.2 Multiple linear regression (MLR)

MLR modeling was performed with the stepwise-selected effects by adjusting the coefficients of the following type of equation:

$$\hat{y} = b_0 + \sum b_i x_i + \sum b_{ii} x_i^2 + \sum b_{ij} x_i x_j + b_{CO2} + \sum b_{iCO2} x_i + e$$

where $\hat{y}$ is the predicted value of the biological variable, b_0 the intercept and e the residual. Continuous environmental variables are designated by x_i or x_j, and their linear, quadratic and interaction coefficients are pointed out as b_i, b_{ii} and b_{ij}, respectively. The coefficients related to CO_2 concentration (ordinal variable) enable to characterize the modification of $\hat{y}$ while switching from the lowest modality (0.035% CO_2) to the highest one (1.5% CO_2). b_{CO2} is for the single effect of CO_2 concentration, whereas b_{iCO2} designates the interactions of CO_2 with continuous environmental variables.

The goodness of fit of the models was assessed by calculating the coefficients of multiple determination (R^2 and R^2 adjusted) and the fitting root-mean-squared error ($RMSE_F$) as follow:

$$R^2 = \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2}$$

$$R^2 \text{ adjusted} = 1 - \frac{\sum (y_i - \hat{y}_i)^2 / (n - k - 1)}{\sum (y_i - \bar{y})^2 / (n - 1)}$$

$$RMSE_F = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n - k - 1}}$$

where n and k are the number of observations and coefficients (apart from b_0) within the model, respectively, y_i and $\hat{y}_i$ are the observed and predicted values of the biological variable, respectively, and $\bar{y}$ is the experimental mean value of the biological variable. The average scale of each biological variable was calculated as follow and exhibited in parallel to $RMSE_F$ as a reference to assess the extent of the deviations:

$$\text{average scale} = \bar{y} - y_{MIN}$$

where $\bar{y}$ and y_{MIN} are the mean and minimal experimental values of the biological variable, respectively.

The statistical significance of the models was assessed by calculating whole-model ANOVA tests with the following expression for the F-ratio:

$$F\text{-ratio} = \frac{\sum (\hat{y}_i - \bar{y})^2 / k}{\sum (y_i - \hat{y}_i)^2 / (n - k - 1)}$$

where the terms are the same than those described for R^2, R^2 adjusted and $RMSE_F$ (cutoff for statistical significance: $p \leq 0.05$).

The importance and statistical significance of each individual effect of the environmental variables were assessed by calculating the related β-weights (= standardized regression coefficients) and ANOVA tests, respectively. For the latter tests, the F-ratio was calculated as follow:

$$R_{RRRRRR} = \frac{\sum (R_R - \mathbb{R}_{R(RR\,R)})^2 - \sum (R_R - \mathbb{R}_R)^2}{\sum (R_R - \mathbb{R}_R)^2 / (R - R - 1)}$$

where the terms with an "*i*" subscript, *n* and *k* have the same significance than described above whereas $\hat{y}_{i\,(k\text{-}1)}$ points out the predicted values of the biological variable in a hypothetical model deprived of the effect (cutoff for statistical significance: $p \leq 0.05$).

Lack-of-fit ANOVA tests were carried out to assess whether the models were lacking one or several major explanatory effect(s) (cutoff for statistical significance: $p \leq 0.05$). The *F*-ratio was calculated as the quotient between the mean square for lack-of-fit error and the mean square for pure error (for details see (SAS, 2012)).

2.8.8 Model cross-validation

Models were cross-validated by the *k*-fold method (k=4) with the Statistica 10 software (StatSoft) by using the data subsets defined in Additional file 2. The goodness of fit of the training models was assessed by calculating R^2, R^2 adjusted and $RMSE_F$ (read above for details about calculations). The deviation of each validation data set from its corresponding training model was assessed in terms of cross-validation root-mean-squared error ($RMSE_{CV}$), which was calculated as follow:

$$RR\,RR_{RR} = R\frac{\overline{\sum (R_R - \mathbb{R}_R)^2}}{R}$$

where y_v are the observed values for the validation data set, $\hat{y}_v$ the values predicted by the training model for the validation data set and v is the number of observations in the validation data set.

2.8.9 Analysis of covariance (ANCOVA)

Analysis of covariance (ANCOVA) was performed through MLR on the basis of standardized data (i.e. data scaled to a mean of 0 and a variance of 1) independently within each cluster. The following general equation was used:

$$\hat{y} = b_0 + \sum b_i x_i + \sum b_{ii} x_i^2 + \sum b_{ij} x_i x_j + b_{CO2} + \sum b_{iCO2} x_i + e$$
$$+ \sum \boldsymbol{b_m} + \sum \boldsymbol{b_{mi}\, x_i} + \sum \boldsymbol{b_{mCO2}}$$

in which the identity of the biological variables is a categorical predictor with *n* modalities ($m_1, m_2, \ldots, m_n$), b_0 is the intercept and *e* the residual. The terms which are not highlighted in bold concern environmental factors alone (read above the section on MLR). The terms in bold refer to the single effects of the biological variables (b_m) and to the second-order interactions between biological and environmental variables ($b_{mi}\, x_i$ for acetate, light, nitrate and ammonium; b_{mCO2} for CO_2). ANCOVA models were characterized by the same goodness of fit and statistical parameters than described above for MLR.

3. Results

A step-by-step overview of the methodology and results is presented in Figure 1.

3.1 Characterization of the environmental regulation of proteins and other biological variables through multivariate statistics

3.1.1 Environmental variables and design of experiments

Environmental variables are light intensity and acetate, CO_2, nitrate and ammonium concentrations in the culture medium (five variables in total). Their characteristics are summarized in Table 2 as the type of each variable, its working range (or modalities for CO_2 concentration) and its unit (similar features as described in (Gérin et al., 2014)).

A statistical design of experiments (DOE) was built to define discrete combinations of the environmental variables to be tested experimentally. Linear effects, quadratic effects and second-order interactions of the environmental variables were considered. The features of this DOE are similar to the design of our previous modeling study (Gérin et al., 2014): the DOE consists of a two level fractional factorial design added with center points and supplementary points found in the Box-Behnken and central composite types of designs. It contains 32 culture conditions, that were already found in the previous DOE (see (Gérin et al., 2014) for more information). Table 3 displays a complete list of DOE items, and a 3D representative example of space covering by environmental variables for light, acetate and nitrate can be found in Additional file 2.

The environmental variables were tested for collinearity by calculating the Pearson's correlation coefficients (R) between them. No statistically significant correlation could be detected ($R \leq 0.16$ with $p \geq 0.3908$), indicating that the design space was uniformly covered.

3.1.2 Biological variables

3.1.2.1 2D-DIGE

The fluorescence image of the internal standard in the Master 2D-gel is presented in Figure 2. In order to minimize the experimental error of protein abundance estimation, the volume exclusion filter for the detection of protein spots was fixed at a tenfold-higher value than recommended by the manufacturer (GE Healthcare). This procedure led to detect 254 spots that could be matched among all 2D-gels (these spots are encircled in Figure 2). Among them, 135 could be identified by mass spectrometry. These spots of interest were defined as continuous biological variables for subsequent statistical analyses (Table 2). They are highlighted in yellow in Figure 2 and complete descriptions about them are provided in Table 4. For each of these spots and each DOE culture condition, the abundance value normalized by the internal standard value was extracted from the Appearance Table of the DeCyder 7.0 BVA module and considered as raw data for subsequent statistical analyses (Additional file 2). To facilitate data treatment and result description, we decided to designate protein spots by their Master number (i.e. their identifier in the Master 2D-gel) followed by their standard name in databases (as found in Table 4).

3.1.2.2 Additional assays

The cellular contents of triglycerides and of some fatty acids and pigments were also considered as continuous biological variables, as well as the respiratory and photosynthetic activities previously reported through DOE approach (Gérin et al., 2014). The respective units of these variables and the analytical methods employed to quantify them are summarized in Table 2 (for more details, read the Material and Methods section). A complete list of the experimental values is provided in Additional file 2.

3.1.3 General features of the data sets

As shown in Additional file 2, one value of protein abundance is available for each protein spot and each culture condition of the DOE (no missing value in the protein data set). Concerning additional assays, there is one missing value for palmitic, stearic and linolenic acids (item 42 of the DOE), two missing values for γ-linolenic acid (items 9 and 42 of the DOE) and three missing values for oleic acid (items 9, 31 and 42 of the DOE). The other additional assays have no missing values. DOE item 42, which is the most frequent missing value among the additional assays, is a center point of the DOE for which two identical measurements exist (items 4 and 41). As explained in the Material and Methods section, all statistical analyses were performed without imputation of the missing values except hierarchical clustering.

3.1.3 Screening of the biological variables influenced by one or several environmental variables

An initial screening was carried out to highlight the protein spots and additional assays relevantly influenced by the environmental variables. This first selection was performed to point out the biological variables for which no further analysis of regulation was necessary (i.e. those for which there was no significant influence of light, carbon or nitrogen). This screening was performed through a methodology coupling partial least squares regression (PLSR) and multiple linear regression (MLR) (the reader is invited to refer to the Material and Methods section for an extensive description of the procedure).

The screening led to the selection of 83 protein spots over 135 (61%) and 15 additional assays over 19 (79%). The results of the screening are described in Additional file 3 for protein spots and in Additional file 4 for additional assays (protein spots of interest are pointed out by orange arrows surrounded by Master numbers in Figure 2 and are presented in the upper part of Table 4 in bold characters).

In numerous cases, two or more spots in 2D-gels were identified as the same protein (91 spots corresponding to 31 different proteins, see Table 4). These observations are due to post-translational modifications generating slight modifications of the isoelectric point (e.g. phosphorylations, deamidations, oxidations) and molecular weight (e.g. complex glycosylations, differences in N- and C-terminal processing) (Nield et al., 2004;Mathy and Sluse, 2008). Only two proteins with multi-identification, FBA3 (three spots) and MDH1 (two spots), did not pass the initial screening for any spot. Less than half of the spots were selected for ATPA (two over six), BLD10 (one over four) and PRK1 (one over three). For the remaining 26 proteins, there were at least 50% of the spots which passed the screening (Additional file 3).

Interestingly, the proteins that are not passing the selection procedure seem to exhibit specificities with regard to their sub-cellular localization or biological function (Table 4 and Additional file 3). The results obtained for the subunits of ATP synthase indicate that the importance of their regulation by light, carbon and nitrogen could mostly depend on their respective cellular compartments: most of the mitochondrial subunits (three over four spots including ASA1, ASA2 and ATP2) passed the screening whereas the chloroplastic and vacuolar subunits were globally rejected (this rejection concerns five spots over eight including ATPA and ATPvA1). A gene set enrichment analysis was carried out in order to verify this qualitative observation (Additional file 5). The "GO cellular component complete" annotation data set of the PANTHER database, in which genes and proteins are classified according to their sub-cellular location, was used. The analysis was carried out by comparing the ATP synthase items found in the unselected protein data set against a reference list consisting of all identified ATP synthase subunits (see Table 4). As shown in Additional file 5, cellular component groups referring to chloroplastic locations exhibited the highest fold enrichment (two folds, positive) with the lowest *p*-value (0.25 without Bonferroni correction for multiple testing). Most of the vacuolar and cell periphery classes were also characterized by a two folds positive enrichment (p=0.437) (both locations are known to specifically contain V-type ATPases, as opposed to the F-type ATP synthases found in chloroplasts and mitochondria). The lowest fold enrichment (more than five folds, negative) was observed for cellular component groups referring to mitochondrial locations (p=0.562). These features tend to support the observations mentioned earlier, but should nevertheless be considered cautiously with regards to the lack of statistical robustness since none of the *p*-values was significant (p>0.05). This is probably due to the very low number of distinct proteins used to perform the enrichment analysis (only four mapped protein IDs in the reference list and two within the list of unselected ATP synthase subunits) (Additional file 5). Moreover, a feature of the technique is to treat each protein as a unique entry (i.e. a unique protein ID) without consideration of the number of spots that are found in each list.

For glycolytic enzymes and cytoskeleton and flagellar constituents, the weakness of the environmental regulation rather seems to be related to the metabolic role of the proteins, without apparent influence of their sub-cellular localization. Among the eight protein spots identified as cytoskeleton and flagellar constituents (basal body protein BLD10, tubulins α and β, actin IDA5), most appear not to be relevantly influenced by the environmental variables (only one BLD10 spot over four passed the screening). None of the glycolytic enzymes (five protein spots including FBA3, PGM1b and PYK1) were selected through the applied procedure. These results suggest that the capacities of glycolysis as well as chloroplastic and vacuolar ATP synthesis were possibly not much influenced by the overall changes of light, carbon and inorganic nitrogen applied in the present study. This is the same for the composition of the cytoskeleton.

3.1.4 Detection and characterization of discrete groups of co-regulated biological variables

3.1.4.1 Partitioning of protein spots through hierarchical clustering

Hierarchical clustering was performed to partition protein spots according to the similarities of their abundance pattern among the culture conditions defined in the DOE.

Results are presented in Figure 3 as a dendrogram with a color range (from green to red) illustrating protein abundance in the different DOE conditions. A two-dimensional distance

plot is also displayed to facilitate cluster visualization. Eight protein spot clusters can be defined according to the general abundance pattern among the tested conditions (Figure 3). For 28 proteins over 31 with multi-identifications, the different spots were clustered together (partitioning among two distinct clusters only for BCR1, ATPA and FNR1). The protein function(s) and cellular location(s) within each cluster were searched in the Pathway Tools section of the ChlamyCyc database (May et al., 2009) and summarized in Table 5.

Most proteins found in cluster 4 are involved in pathways related to acetate assimilation (e.g. acetyl-CoA synthesis, glyoxylate cycle, TCA cycle and gluconeogenesis) (Figure 3, Table 5). Acetyl-CoA is generated from acetate by acetyl-CoA synthetase and is then metabolized through the glyoxylate and TCA cycles. The reducing equivalents and C_4 intermediates produced by these pathways can then be directed to gluconeogenesis, as previously shown in *C. reinhardtii* (Johnson and Alric, 2012). Aspartate aminotransferase (involved in anaplerosis and reductant transport) and catalase (participating to ROS detoxification) were also partitioned in the same cluster. This observation might be related to the higher electron input possibly induced by acetate assimilation. Such a feature could heighten the intracellular redox state and the ROS production rate, and make necessary to develop higher capacities of reductant transport and ROS detoxification.

In cluster 7, most proteins are related to bioenergetic processes: enzymes of the TCA cycle, components of the mitochondrial ATP synthase, coproporphyrinogen III oxidase (precursor of heme and chlorophyll) and proteins involved in cell redox signaling (glutathione-S-transferase and 14-3-3 protein FTT2) (Foyer and Noctor, 2003;Roberts, 2003). Argininosuccinate synthase (which catalyzes the last, irreversible reaction of arginine biosynthesis) can also be found in this cluster. The carbon skeletons, reducing equivalents and ATP molecules generated by mitochondrial catabolism are important substrates for amino acid biosynthesis, as extensively reported (Turpin, 1991;Huppe and Turpin, 1994;Foyer et al., 2011). With this regard, the partitioning of biological variables related to mitochondrial catabolism and amino acid biosynthesis into the same cluster might reflect the need to coordinately regulate the capacity of both groups of pathways.

Most members of cluster 8 are chloroplastic proteins involved in the assembly, the architecture and/or the stabilization of core photosystems and light-harvesting antennae. Ferredoxin-NADP reductase, an enzyme participating to the photosynthetic electron transport, can also be found there (two spots in cluster 7 and two spots in cluster 8). In cluster 5, most proteins are related to protein biosynthesis, maturation, stabilization, targeting and/or assembly into complex structures. In clusters 1 and 6, proteins are all involved in anabolic processes (especially the Calvin cycle). Finally, cluster 2 can be described as a tote-bag in which there is no clear tendency with regard to the general function of proteins.

Gene set enrichment analyses of pathways were carried out to verify the co-segregation of proteins participating to common metabolic functions. The analyses were performed in the "GO Pathways" annotation data set of the PANTHER database. The proteins within each cluster were compared to a reference list made of all proteins used to perform hierarchical clustering (47 different proteins). The pathways exhibiting a positive enrichment comparatively to the reference list are displayed in Additional file 6 for each cluster. Among the 47 proteins in the reference list, 45 could be mapped to at least one pathway entry in the database. Unfortunately, for 28 of these proteins, this entry was the unclassified category. In clusters 1, 2, 3, 6 and 8, the unclassified category contained nearly all proteins and showed a positive enrichment of about 1.5 fold (except in cluster 6 for which the enrichment was only

worth 1.15). In cluster 4 (three unclassified proteins over seven), acetate utilization and asparagine/aspartate biosynthesis exhibited the highest fold enrichment (6.43) (Additional file 6). In cluster 5 (four unclassified proteins over seven), a five folds enrichment was observed for two pathways related to amino acid biosynthesis: glutamine/glutamate conversion and S-adenosylmethionine biosynthesis. In cluster 7 (five unclassified proteins over ten), pathways related to bioenergetics (heme biosynthesis and pyruvate metabolism) and amino acid biosynthesis (leucine and arginine) showed the highest fold enrichment (4.50). The same result was also obtained for two redox signaling pathways as well as for the degradation of ascorbate (which is a powerful antioxidant) (Hüttemann et al., 2007;Smirnoff, 2011;Lamb et al., 2015). These results collected for clusters 4, 5 and 7 are in agreement with the considerations mentioned earlier with regard to the metabolic function of proteins. It should be noticed that nearly all *p*-values of the enrichment analysis are unsignificant ($p>0.05$; see Additional file 6); results should therefore be considered cautiously due to the lack of statistical robustness. The reasons for that could be the same as those described in section 3.1.3.

3.1.4.2 Integration of the additional assays within specific protein clusters

Hierarchical clustering was reiterated by also including additional assays in the analysis, together with protein spots. The purpose of that was to partition the assays in the different protein clusters according to pattern similarities among the DOE conditions. Results are summarized in Figure 3.

All respiratory parameters (CR, MA_{CYT} and MA_{ALT}) and fatty acids (palmitic, oleic and linolenic acids) are associated with protein cluster 7. Palmitic, oleic and linolenic acids constitute highly-energetic substrates for β-oxidation in the mitochondrion, and are known to mediate the activity of the mitochondrial uncoupling proteins (Jezek et al., 1998). Moreover, linolenic acid is the most abundant fatty acid found in plant thylakoid membranes (Murphy, 1986). β-carotene and violaxanthin also exhibit pattern similarities with protein cluster 7. They are the only carotenoids for which the biological function cannot be substituted by other pigments in case of mutational deletion, and are thought to protect the photosynthetic apparatus from photo-oxidative damages (Trebst, 2003). The remaining pigments (chlorophylls *a* and *b*, total carotenoids, neoxanthin, lutein) were rather partitioned with protein cluster 8. For $\varphi PSII_{800}$ and P_{800}, the dependence upon DOE conditions is related to protein cluster 5.

The term "cluster" will be used thereafter to designate each group of biological variables (proteins and assays) exhibiting a similar pattern among the DOE conditions, as shown by hierarchical clustering.

3.1.4.3 Characterization of the cluster-specific attributes by multivariate analysis

A principal component analysis (PCA) was performed with all biological variables in order to characterize cluster-specific regulatory tendencies with regard to the particularities of the DOE conditions. Figure 4A shows the results of the PCA as the corresponding score plot and loading plot, based on the two first principal components. The first and the second components account for 29.7 and 24.1% of the variability, respectively, with only 9.6% for the third component (see Additional file 7).

In the loading plot (bottom-right of Figure 4A), the vectors of the biological variables are colored according to their respective cluster. As expected, biological variables within each cluster appear to be grouped together as vector bundles pointing toward a specific direction. As illustrated in Figure 4A, the correlation among biological variables is the highest within clusters 3, 6 and 8, as evidenced by the narrow angle covered by their vector bundles. In the other clusters however, the observation of a much more important angle (close to 90°, with a maximal amplitude for cluster 5) indicates that the correlation between some biological variables can be very weak despite the high correlation between neighboring vectors (the correlation matrix of each cluster is displayed in Additional file 8). For example, in cluster 5, a correlation of 0.93 (highest value within the cluster) is observed between two isoforms of UPTG1 (spots 112 and 215) but the correlation is of only -0.18 between 4-HSP70B and 3-BLD10.

In order to assign the cluster-specific grouping of biological variables to specificities of the DOE culture conditions, the score plot was reproduced in five identical copies (one per environmental variable) and each observation was marked with its respective DOE value using a specific color scale (Figure 4A). A shown there, the observations can be divided into five groups according to their relative position to the first and the second principal components. The distinction between the two groups of the inferior quadrants is essentially due to the third principal component (data not shown). Interestingly, each group of observations exhibits specific tendencies regarding the value of one or several environmental variable(s). Figure 4B summarizes the tendencies observed within each quadrant and also displays angular covering by the vectors of each cluster under the form of a biplot-like scheme.

In addition, in-cluster PLSRs were carried out as a supplementary way to assess regulatory specificities. Variable importance in projection (VIP) of each environmental variable is displayed in Figure 4C for the different clusters (see Additional file 7 for details about PLSR results). For VIPs exceeding the cutoff value of 1, the sign of the coefficient in the PLSR models is also provided.

As highlighted in Figure 4B and C, negatively correlated clusters (cluster 6 vs 8; cluster 4 vs 5; clusters 1-2-3 vs 7) exhibit an opposite regulation by specific environmental variables. Biological variables found in clusters 6 and 8 are both controlled by acetate concentration and light intensity with a positive influence of these factors in cluster 6 and a negative one in cluster 8. For the members of clusters 4 and 5, there is a substantial effect of light intensity which appears to be negative in cluster 4 and positive in cluster 5. Finally, nitrate and/or ammonium concentrations seem to be the most important factors regulating biological variables in clusters 1, 2, 3 and 7, with a positive influence in clusters 1-2-3 and a negative one in cluster 7. It is worth noticing that neither PCA nor PLSR enable the visualization of more complex effects than linear ones. Results presented in Figure 4B and C are therefore likely to provide an incomplete overview of in-cluster regulatory tendencies.

Altogether, PCA and PLSR results indicate that regulatory tendencies exist within each cluster with some in-cluster subtle regulatory divergences, as suggested by the observation of a quite weak correlation among some biological variables. These divergences were further characterized by an analysis of covariance (ANCOVA) through MLR (one model per cluster), by introducing the identity of the biological variables as a categorical model predictor (for details, read the Material and Methods section). ANCOVA results are presented in Additional file 9. All models are significant (p<0.0001) with relatively high values of R^2 adjusted (0.62

on average) and low fitting root mean squared errors ($RMSE_F \leq 33\%$ of the response average scale). For each individual biological variable, the effects of the different environmental factors were statistically compared to the overall regulation within the cluster, and the significant differences ($p \leq 0.05$) were further characterized by their respective β-weights (standardized regression coefficients). This approach enabled on the one hand to identify the biological variables exhibiting an outlying regulation within each cluster, and on the other hand to determine which environmental variables were responsible for this divergence (see the summary scheme in Additional file 9). Consistently with PCA results, ANCOVA demonstrates that the environmental regulation is quite homogenous in clusters 3, 6 and 8 but shows a more important proportion of discrete divergences in the other clusters.

3.1.5 Independent modeling of the dependence of each biological variable upon light, carbon and nitrogen

The influence of the environmental factors (Table 2) was modeled independently for each biological variable through MLR. Single effects and second-order interactions between environmental variables were considered, as well as linear and quadratic effects for the continuous ones (for details, read the Material and Methods section). Prior to MLR modeling, stepwise regression was carried out with minimum AICc (corrected Akaike information criterion) as stopping rule in order to reduce the number of coefficients and limit the probability of overfitting (Gérin et al., 2014).

3.1.5.1 Model parameters and cross-validation

The model parameters and the regression equation of each biological variable are provided in Additional file 10. The values predicted by this equation for the different DOE culture conditions are listed in Additional file 11. On average for all biological variables, $RMSE_F$ is worth 40% of the response average scale with a standard deviation of 11%, and the mean R^2 adjusted is equal to 0.61 with a standard deviation of 15%. The whole-model ANOVA *p*-values are statistically significant for every biological variable, and the lack-of-fit is significant for 101-PRK1 only ($p=0.0047$). These results indicate that the effects included in the models are likely to be sufficient to explain most of the variability of the biological variables.

Models were cross-validated by the *k*-fold method with $k=4$ in order to spot potential overfitting. Data subsets are described in Additional file 11. Cross-validation results are displayed in Additional file 12 as superimposed bar charts enabling comparison between the cross-validation root-mean-squared error ($RMSE_{CV}$) and both the training $RMSE_F$ and the response average scale (references to assess the extent of $RMSE_{CV}$). On average, $RMSE_{CV}$ is worth 51% of the response average scale with a standard deviation of 15%. $RMSE_{CV}$ exceeds the training $RMSE_F$ by 32% on average (=12% of the response average scale) with a standard deviation of 19% (=7% of the response average scale). Since $RMSE_{CV}$ is not harshly above $RMSE_F$, models are likely not to overfit for most biological variables. As shown in bar charts (Additional file 12), exceptions to this assumption could be 193-GAP, 108-CYN38, 182-FNR1, 29-CPN60A and 71-EEF1A ($RMSE_{CV}$ exceeds $RMSE_F$ by more than two-thirds, suggesting that the less significant factors - with $0.01<p<0.05$ - should be considered cautiously).

3.1.5.2 Relative importance and mathematical profile of the environmental variables

Figure 5 illustrates the β-weights associated to the statistically significant coefficients ($p \leq 0.05$) as a green-to-red heat map. Protein spots and additional assays are sorted by cluster, and the empty cells are either for insignificant or stepwise-unselected effects. A complete list of β-weights and *p*-values is provided in Additional file 10.

As illustrated in Figure 5, biological variables are mostly regulated through single, linear effects of the environmental factors (over 50% of significant linear coefficients for the continuous variables and 26% for CO_2 concentration). Only 16% of the quadratic coefficients are globally significant but this proportion reaches 29% for nitrate concentration. The second-order interaction between nitrate and ammonium is statistically significant for 48% of the biological variables, with only 2-16% for the other interactions. This observation suggests that the nature of the inorganic nitrogen source and the balance between its different molecular species are likely to be key regulators of cellular metabolism.

Nearly all biological variables are nonetheless regulated through complex superimpositions of linear effects, quadratic effects and/or second-order interactions of the environmental variables (Figure 5). That renders the visualization of regulation quite difficult and makes necessary to perform a case-by-case mathematical simulation for every biological variable and environmental factor. For technical reasons, it is not possible to present such numerous simulations here. Consequently we rather chose to build generalized simulation plots considering every possible situation (Figure 6) as a key to read the results presented in Figure 5.

In good agreement with PCA and ANCOVA, the identity, sign and relative importance of the significant coefficients are especially homogenous among biological variables in clusters 3 and 8 (Figure 5). As expected, the regulatory differences among the clusters appear to be far more remarkable than within each individual one (existence of a clear regulatory pattern unique to every cluster). The results obtained by PCA and PLSR with regard to the in-cluster regulatory specificities (Figure 4B and C) are consistent with MLR results (Figure 5).

3.2 Description of the environmental regulation of the biological variables according to their metabolic function

3.2.1 Biological variables related to photosynthesis and protein metabolism

Most components of light-harvesting antennae (LHC proteins and pigments, cluster 8) are controlled by light intensity and acetate concentration through negative linear effects (Figure 5, Table 5). Consistently, acclimation to increasing irradiance has long been known to involve a substantial down-regulation of LHC proteins and to lower the cellular pigment amount (Neale and Melis, 1986;Falkowski and LaRoche, 1991;Teramoto et al., 2002;Durnford et al., 2003;Nield et al., 2004). Moreover, the contents of chlorophyll *a* and *b* have been reported to be decreased in *C. reinhardtii* cells grown in the presence of acetate (Boyle and Morgan, 2009). On the contrary, most of the actors of the enzymatic machinery involved in protein synthesis and maturation (clusters 5 and 6) are regulated by light intensity and acetate concentration through positive linear effects (with the exception of EEF1A for which there is no influence of light, see Figure 5). Accordingly, the cytosolic heat-shock protein HSP70A has already been reported to be induced by light at the mRNA abundance level (Von Gromoff et al., 1989).

In addition, most components of the light-harvesting antennae (cluster 8) tend to be regulated by nitrate concentration through a quadratic convex profile (minimum estimated around 12.5 mM) (Figure 5, Figure 6, Additional file 10). Interestingly, a reciprocal concave control is exerted by this factor on a series of other photosynthesis-related biological variables: P_{800} (the maximal gross O_2 evolution) and two enzymes catalyzing regulatory steps of the Calvin cycle (sedoheptulose-1,7-bisphosphatase and phosphoribulokinase) (Hahn et al., 1998;Raines, 2003). In some circumstances, P_{800} can be regarded as an indicator of the capacity of the metabolic pathways consuming the photo-generated NADPH and ATP, such as the Calvin cycle (Badger et al., 2000).

Remarkably, the enzymes involved in protein synthesis and maturation (cluster 5) are the only group of biological variables exhibiting a clear regulatory tendency upon CO_2 concentration (positive effect) (Figure 5, Table 5).

3.2.2 Biological variables related to the Calvin cycle

The regulation of the Calvin cycle enzymes (clusters 1 and 6) by light, carbon and nitrogen appears to be quite heterogeneous (Figure 5, Table 5). Transketolase is mostly regulated through linear effects of nitrate concentration. Rubisco large subunit, sedoheptulose-1,7-bisphosphatase and phosphoribulokinase are controlled by quadratic effects of nitrate concentration with an additional quadratic influence of acetate concentration for RubisCO large subunit and phosphoribulokinase. A linear effect of light intensity is also observed for the latter enzyme. The abundance of glyceraldehyde-3-phosphate dehydrogenase mostly depends on linear effects of acetate concentration, nitrate concentration and light intensity. No significant effect of CO_2 concentration could be detected for any of these enzymes, except a second-order interaction between acetate and CO_2 concentrations for some spots of RubisCO large subunit (Figure 5). Accordingly, CO_2 concentration has already been reported to exert no relevant influence on the cellular abundance of the latter protein (Borkhsenious et al., 1998;Mitchell et al., 2014).

3.2.3 Biological variables related to acetate assimilation

Nearly all proteins involved in acetate assimilation (cluster 3) are controlled by light intensity through negative linear effects. This concerns acetyl-CoA synthetase as well as enzymes of the glyoxylate cycle, TCA cycle and gluconeogenesis (Figure 5, Table 5). In line with these observations, light has been shown to act as a negative regulator of the glyoxylate cycle in plants (Allen et al., 1988), and lower mRNA levels have been reported for isocitrate lyase consequently to light exposure in *C. reinhardtii* (Petridou et al., 1997).

Interestingly, acetate concentration alone does not appear to significantly influence the machinery responsible for its own assimilation at the protein abundance level (Figure 5). This is consistent with the observation that the genetic expression of malate synthase and isocitrate lyase occurs in both the presence and the absence of acetate in plants (Graham et al., 1994). A negative interaction between acetate concentration and light intensity could nonetheless be detected for most proteins involved in acetate assimilation. Therefore, increasing acetate availability is expected to strengthen the negative influence exerted by light intensity on the acetate assimilatory machinery, so that the most important effect of light will be observed in case of high acetate availability (Figure 5, Figure 6). A positive second-order interaction between acetate and CO_2 concentrations was also detected for a few proteins participating to

acetate assimilation (phosphoenolpyruvate carboxykinase and aconitase) (Figure 5). This possibly indicates that the total carbon availability could contribute to control acetate assimilation to some extent.

Acetate assimilatory enzymes also tend to be regulated by nitrate and ammonium concentrations through positive linear effects (less well defined tendency in comparison to light) (Figure 5). This suggests that the abundance of these proteins could be controlled by the total availability of inorganic nitrogen. Accordingly, the mRNA levels of aconitase and phosphoenolpyruvate carboxykinase have been reported to be very sensitive to nitrogen deprivation in *C. reinhardtii* (Miller et al., 2010).

3.2.4 Biological variables related to bioenergetic processes

Nearly all the biological variables involved in bioenergetic processes found in cluster 7 are regulated by nitrate and ammonium concentrations through negative linear effects (Figure 5, Table 5). Such as for acetate assimilatory enzymes, this could indicate that the mitochondrial and chloroplastic bioenergetic pathways are very sensitive to the total availability of inorganic nitrogen at the protein abundance level. This assumption is strengthened by the observation of a positive second-order interaction between nitrate and ammonium concentrations (Figure 5). Such an interaction could contribute to improve the tightness of the metabolic response, by attenuating the negative influence of each factor when the availability of the other nitrogen source increases in the medium.

3.2.5 GroEL-homolog chaperonin CPN60A

The abundance of this protein (the only one in cluster 3) is strongly regulated by ammonium concentration through positive linear effects (Figure 5). This observation might be related a possible role of CPN60A in the enhancement of the stability of the enzymatic machinery for photosynthesis and nitrate reduction, as reported in cyanobacteria in case of thermal stress (Rajaram and Apte, 2008). The observation of a negative second-order interaction with CO_2 concentration also indicates that the influence of ammonium might be attenuated in high CO_2-grown cells. This effect of CO_2 could be related to the well-known participation of CPN60A to the assembly of RubisCO holoenzyme in plants (see (Hauser et al., 2015) for review).

4. Discussion

The present work is focused on studying the influence of simultaneous variations of light, carbon and inorganic nitrogen on the cellular proteome of *C. reinhardtii*. For this purpose, design of experiments (DOE) and sequential multivariate analyses were used to model protein regulation upon overall environmental changes. Proteomic results were completed by additional assays for respiration, photosynthesis, and cellular contents of some lipids and pigments, and the data of these assays were integrated into proteomic results through multivariate statistics. To date, most reported efforts have been focused on studying the effects of one or two environmental variables on photosynthetic metabolism (keeping the other variables constant). Moreover, little information was available in literature concerning the mathematical influence profile of each variable and its relative weight.

Over the last decade, a very wide panel of omics-based approaches has been developed to gain deeper understanding of many aspects of cellular biology. With regard to the huge amount of data generated by these techniques, efficient bioinformatics methods of meta-analyses have been developed to reconstitute biological systems. In this context, dealing with data heterogeneity is the key problem (Fukushima et al., 2009;Mochida and Shinozaki, 2011). The use of multivariate statistical approaches could help solving this problem to some extent, by making possible to perform an overall regulation study with a single experimental design.

4.1 The present set of sequential multivariate analyses is suitable for the characterization of the environmental regulation of *C. reinhardtii* metabolism

As already described in details, the results of the regression-based initial screening are homogenous for the different spots of proteins with multi-identifications (Additional file 3). In the individual MLR models obtained for the selected biological variables, an important proportion of the variability can be explained by light, carbon and nitrogen (Figure 5, Additional file 10). These elements indicate that the screening procedure that we used here is reliable for the present data set.

Hierarchical clustering is a key element of the present work that enabled to partition biological variables according to their regulatory similarities. Such a methodology had already been employed by Höhner et al. for the analysis of *C. reinhardtii* proteomic data to study the response to environmental changes (iron availability and trophic status) (Höhner et al., 2013). The authors demonstrated that the proteins participating to a common biological function tended to be grouped together. Similarly here, hierarchical clustering enabled to partition biological variables into eight co-regulated clusters corresponding to specific biological processes: Calvin cycle (cluster 1), acetate assimilation (cluster 4), protein synthesis and maturation (cluster 5), anabolic pathways (cluster 6), processes of energy transduction (cluster 7) and composition of the photosynthetic apparatus (cluster 8) (Figure 3, Table 5). The observation of such a weak number of clusters is outstanding with regard to the diversity of the environmental perturbations applied here.

An overview of the regulation by light, carbon and nitrogen within each cluster was further characterized by PCA, PLSR and ANCOVA (Figure 4, Additional file 9). These analyses indicated the existence of slight in-cluster differences with regard to the influence of the environmental variables. These observations were particularly marked for clusters 1, 2, 4, 5 and 7, and suggested that subtle regulatory divergences could exist within each cluster despite the existence of a common pattern. These divergences among biological variables were therefore assessed by modeling the influence of light, carbon and inorganic nitrogen through MLR, independently for each protein spot and additional assay. In contrast with PCA and PLSR, these analyses enabled to simulate the mathematical influence profile of each environmental variable by taking into account quadratic effects and second-order interactions (Figure 5, Figure 6). As expected, the differences were much less marked within the clusters than among them, confirming the existence of a clear regulatory pattern unique to every cluster.

4.2 The present analyses provide deeper insight into the metabolic adaptations set up in response to overall environmental changes

4.2.1 Light, carbon and inorganic nitrogen exert no influence on a series of biological variables associated to specific sub-cellular compartments or biological functions

According to the results of the initial regression-based screening, most proteins which are not substantially influenced by light, carbon or nitrogen (Table 4, Additional file 3) seem to belong to discrete sub-cellular compartments or functional groups. On the one hand, as verified by gene set enrichment analysis, this absence of environmental regulation concerns the chloroplastic and vacuolar subunits of ATP synthase. On the other hand, no incidence of light, carbon and nitrogen could be noticed for the glycolytic enzymes nor for the cytoskeleton and flagellar components analyzed here, independently of their sub-cellular localization. Previous studies indicated that light might influence glycolysis by inhibiting pyruvate kinase in *C. reinhardtii* (Xue et al., 1996). As suggested here, this possible light-mediated inhibition of glycolytic activity might not be associated to a significant decrease of the capacity of the pathway. In *C. reinhardtii*, some subunits of the chloroplastic and vacuolar ATP synthases are also known to be regulated by light through the thioredoxin system (Lemaire et al., 2004). This variation of activity does not seem to correlate with a significant modification of protein abundance.

Among the functional assays for respiration and photosynthesis, only NPQ_{800} did not pass the initial screening (Additional file 4). This may be related to the lower ability of *C. reinhardtii* to set up non-photochemical quenching of chlorophyll fluorescence in comparison with plants (Finazzi et al., 2006).

4.2.2 Influence of light, carbon and inorganic nitrogen on the cellular metabolism

As shown in Figure 5, the regulation of most selected biological variables occurs through linear effects of light, acetate, nitrate and ammonium. For CO_2 concentration, the number of significant coefficients ($p \leq 0.05$) is twice lower in comparison with the other variables. Moreover, no cluster-specific regulatory tendency can be distinguished regarding this factor, except in cluster 5 in which there is a positive influence of CO_2 for many biological variables. Remarkably, no influence of CO_2 concentration could be detected here for Calvin cycle enzymes, including RubisCO as already reported at the abundance level (Borkhsenious et al., 1998;Mitchell et al., 2014).

We hypothesize that the weakness of CO_2 influence could arise from two particularities of the experimental design. Firstly, the cellular density in algal cultures was relatively weak at the time of harvest (biomass: 250 $\mu g.mL^{-1}$). The uptake of CO_2 by algal cells was therefore probably not limited by the rate of CO_2 diffusion in the aqueous phase. In these conditions, the induction of the carbon-concentrating mechanism (CCM) under 350 ppm CO_2 might have been sufficient to buffer the variations of CO_2 levels in the local environment of RubisCO (Moroney et al., 2011;Wang et al., 2011;Kupriyanova et al., 2013). Accordingly, the acclimation of *C. reinhardtii* cells to low CO_2 has been associated with increased levels of several CCM proteins without modification of the abundance of RubisCO large and small subunits (Mitchell et al., 2014). Secondly, the maximal light intensity used here (200 $\mu mol_{photons}.m^{-2}.s^{-1}$) is not high enough to induce saturation the photosynthetic electron transport chain (Sueltemeyer et al., 1986;White and Critchley, 1999). The production rates of NADPH and ATP (rather than the availability of CO_2) are therefore likely to constitute

limiting factors for the Calvin cycle in the present conditions. Altogether, these different elements might rationalize that huge modifications of CO_2 availability (from 350 ppm to 1.5%) are shown here to induce only slight metabolic adaptations.

The features discussed below regarding the influence of light, carbon and inorganic nitrogen on the cellular metabolism are illustrated in Figure 7. Panel A is for nitrate and ammonium; panel B is for light, acetate and CO_2. These schemes represent interpretations of our results, mostly related to changes in protein abundance.

4.2.2.1 Influence of nitrate and ammonium

As shown in Figure 5, nitrate and ammonium exert a significant influence on biological variables related to carbon metabolism (Calvin cycle, glyoxylate cycle, gluconeogenesis) and processes of energy transduction (respiration, photosynthesis, light harvesting) (see also Table 5). The predominant regulatory nitrogen form and its mathematical influence profile are specific for each pathway: for example, light-harvesting antennae components (cluster 8) are regulated by nitrate concentration through a quadratic influence profile, whereas processes of energy transduction (cluster 7) rather depend on the total availability of inorganic nitrogen through negative effects. To date, the influence of nitrate and ammonium concentrations had poorly been investigated, but dramatic effects of nitrogen deprivation on many aspects of biological functions had nonetheless been reported (Plumley and Schmidt, 1989;Turpin, 1991). Altogether, these data and the present results emphasize that the inorganic nitrogen source is a key factor controlling the energetic balance of the cell. Interestingly, for nearly half of the biological variables, a significant interaction could be detected between nitrate and ammonium concentrations (Figure 5). This suggests that the balance between these two forms also exerts a particularly important control on biological processes, probably because of the higher energetic requirements of nitrate assimilation (Fernandez et al., 2004).

4.2.2.1.1 Nitrate quadratically influences the machineries for light-harvesting, photosynthesis and CO_2 fixation: A way to adjust the photo-production of reductant, ATP and carbon skeletons to the assimilation of this N source?

The components of light-harvesting antennae (LHC proteins ant pigments in cluster 8) are shown here to be regulated by nitrate concentration through a convex profile with an inflexion point around 12.5 mM (Figure 5, Figure 6, Additional file 10). Interestingly, several other biological variables related to photosynthesis are controlled by nitrate through a reciprocal concave profile: Calvin cycle enzymes (sedoheptulose-1,7-bisphosphatase, phosphoribulokinase, some spots of RubisCO large subunit), linolenic acid (the most abundant fatty acid in thylakoid membranes), P_{800} (the gross photosynthetic O_2 evolution) and some spots of ferredoxin-NADP reductase (Figure 5, Figure 6). These data suggest that certain aspects of light harvesting, photosynthetic electron transport and CO_2 fixation are coordinately regulated by nitrate concentration. Remarkably, the observation of quadratic profiles indicates the existence of an optimal concentration of this factor for photosynthesis. The experimental elements are nevertheless insufficient to rationalize the occurrence of two types of reciprocal quadratic effects.

Photosynthesis is an important source of reductant, ATP and carbon skeletons for nitrogen assimilation (Turpin, 1991). Consequently, it can be argued that the regulation of photosynthesis by nitrate concentration arises from the need to adjust the rate of reductant, ATP and carbon skeleton production to the rate of nitrate assimilation. That had already been

suggested for P_{800} in a previous publication (Gérin et al., 2014). The data reported here indicate that nitrate-induced adaptations occur at two levels: (i) at the level of the photosynthetic electron transport chain as a way to control the production of reductant and ATP, and (ii) at the level of the Calvin cycle as a way to adjust the rate of carbon skeleton production (Figure 7A). Such adaptations are likely to contribute to the regulation of the carbon-to-nitrogen balance of the cell under changing nitrate availability in the culture medium.

4.2.2.1.2 Nitrogen down-regulates pathways contributing to its assimilation at the protein abundance level

A recent study reported the proteomic adaptations of algal cells upon changes of ammonium availability in the culture medium (testing of four different concentrations) (Lee et al., 2012). A drastic increase of the abundance of several TCA cycle enzymes (citrate synthase, isocitrate dehydrogenase, α-ketoglutarate dehydrogenase, succinate dehydrogenase, malate dehydrogenase) and of glutamine synthetase could be detected while decreasing ammonium concentration. These results were attributed to the need to heighten the capacity for amino acid biosynthesis through the GS/GOGAT cycle and anabolic pathways (requiring organic acids as carbon skeletons) in case of low nitrogen supply (Lee et al., 2012). Here the observation that total nitrogen availability (nitrate + ammonium) exerts a negative influence on biological variables involved in mitochondrial catabolism (notably citrate synthase and isocitrate dehydrogenase) and amino acid biosynthesis (argininosuccinate synthase) (see cluster 7 in Figure 5) is in agreement with this assumption. In line with that previous study, a strong negative influence of ammonium concentration on the abundance of glutamine synthetase could also be detected here (Figure 5). The present work further demonstrates that nitrogen influence prevails over the effects of light and carbon for the regulation of TCA cycle and amino acid biosynthetic enzymes at the protein abundance level (no clear regulatory tendency upon changes related to light, CO_2 and acetate, see Figure 5 and 7A).

4.2.2.2 Influence of light, acetate and CO_2

4.2.2.2.1 Light-mediated activation of the Calvin cycle does not always correlate to higher protein abundance

Calvin cycle enzymes are known to be activated by light through redox mechanisms mediated by the thioredoxin system. That enables to accelerate the turnover of NADPH and ATP when light intensity increases, with a concomitant improvement of CO_2 fixation (Perchorowicz et al., 1981;Brooks et al., 1988). Remarkably, the data presented here indicate that the thioredoxin activation of Calvin cycle enzymes is not always associated to higher protein abundance levels. Statistically significant coefficients were indeed detected for some enzymes (glyceraldehyde-3-phosphate dehydrogenase, phosphoribulokinase) but in other cases light was not shown to be a regulatory factor (RubisCO large subunit, sedoheptulose-1,7-bisphosphatase, transketolase) (Figure 5). Accordingly, no major changes of the abundance of RubisCO large and small subunits could be detected during the dark-to-light transition in *C. reinhardtii* (Mitchell et al., 2014). Light had previously been reported to considerably enhance the mRNA levels for sedoheptulose-1,7-bisphosphatase in *C. reinhardtii* (Hahn et al., 1998) but our results indicate that this increase in transcript abundance does not result in higher protein amount.

4.2.2.2.2 Adaptation to increasing irradiance heightens the capacity to assembly and protect photosystem II reaction centers

The quantum yield of photosystem II under saturating light ($\varphi PSII_{800}$) was partitioned in the same cluster (n°5) as the components of the machinery for protein synthesis and maturation (Figure 3, Table 5). In this group, biological variables are positively influenced by light, acetate and CO_2 (Figure 5). Interestingly, increasing light irradiance is known to accelerate the turnover of the D1 protein of photosystem II as a way to replace photo-damaged reaction centers (Schuster et al., 1988). In this context, the chloroplastic heat-shock protein 70B has been suggested to participate to both the protection and repair of the reaction centers (Schroda et al., 1999). Here the observation that $\varphi PSII_{800}$ and HSP70B are found in the same light-dependent cluster is in agreement with this postulated role of HSP70B.

$\varphi PSII_{800}$ and P_{800} were partitioned in the same cluster, but nonetheless differ from each other regarding the effects of acetate, nitrate and CO_2 concentrations (Figure 5). These features might be attributable to the fact that P_{800} does not only depend on intrinsic properties of the photosynthetic apparatus, but is also modulated by interactions of photosynthesis with other metabolic pathways (Calvin cycle, photorespiration, Mehler reaction, etc.) (Badger et al., 2000). The molecular mechanisms underlying P_{800} environmental regulation are therefore likely to be more complex than $\varphi PSII_{800}$.

4.2.2.2.3 Heightening the metabolic rate and decreasing the capacities for light and acetate assimilation: A double strategy to limit the harmful effects of excess energy input?

In *C. reinhardtii*, the metabolic rate is known to be stimulated by light, acetate and CO_2 (Sager and Granick, 1953;Yang and Gao, 2003;Boyle and Morgan, 2009). Here data demonstrate that these environmental variables exert a positive influence on the enzymatic machinery for protein synthesis and maturation (Figure 5). That could indicate that the capacity for protein turnover is increased in response to light, acetate and CO_2, possibly as a way to support the higher metabolic rates induced by heightening these variables (Figure 7B).

Conversely, light was shown here to exert a negative influence on some pathways related to carbon assimilation, i.e. acetate metabolism (acetyl-CoA synthetase, glyoxylate cycle, TCA cycle, gluconeogenesis) and light harvesting (indirectly connected to CO_2 fixation through the photo-production of reductant and ATP as substrates of the Calvin cycle) (Figure 5). In addition, a negative influence of acetate concentration could also be detected for light-harvesting antennae components. For acetate assimilatory enzymes, the influence of this factor occurs indirectly though a negative second-order interaction with light intensity (Figure 5). This interaction strengthens the negative influence of light while increasing acetate availability, in such a way that the more important effect of light is observed in case of high acetate concentration (see also Figure 6). Accordingly, cross-talk between light and acetate signaling pathways has already been reported to play a key role in the regulation of malate synthase, a specific enzyme of the glyoxylate cycle (Nogales et al., 2004). However varying acetate concentration alone (i.e. without changing light) appears to be insufficient to induce metabolic adaptations of the acetate assimilatory pathways (Figure 5). This observation that acetate does not exert a direct control on its own assimilation at the protein abundance level is quite remarkable.

The negative influence of light on the capacity of the photosynthetic antennae has long been known to avoid over-reducing the photosynthetic apparatus while increasing irradiance. This adaptation enables to control light energy capture and to prevent the occurrence of oxidative stress within the cell (Falkowski and LaRoche, 1991;Teramoto et al., 2002). By extension, the aforementioned adaptations related to acetate assimilation and light harvesting (Figure 5) could be a way to limit the energy input while increasing the availability of electron sources such as light and acetate. Overall, accelerating the metabolic rate and decreasing the capacities for light and acetate assimilation might be a double strategy enabling to prevent primary metabolism blocking and to limit oxidative damages consequently to increased availabilities of light and acetate (Figure 7B).

5. Conclusions

Altogether, the present results support that the environmental regulation of the primary metabolism is a multifactorial issue, since nearly all biological variables were found to be influenced by complex superimpositions of linear effects, quadratic effects and/or second-order interactions of the environmental variables. That supports the usefulness of studying regulation in a context where light, carbon and nitrogen are varied simultaneously in the medium, in order to guarantee that the observations are not specific of a particular physiological state. The quadratic effects exerted by nitrate concentration on some components of the machineries for photosynthesis and CO_2 fixation appear to us as particularly interesting. In our opinion, this influence of nitrate would deserve to be further investigated with regard to its possible consequences on primary productivity and industrial biomass yields (potential existence of an optimal nitrate concentration). If combined to omics methods exhibiting higher output levels than 2D-DIGE (gel-free proteomics, microarray, etc.), we think that the present statistical methodology could enable to considerably improve current understanding of systems biology in diverse organisms. In this context, extensive sequential statistical analyses could help dealing with heterogeneous experimental and analytical procedures to unveil hidden information in increasingly large biological data sets.

List of abbreviations

2D-DIGE	Two dimensional-differential in-gel electrophoresis
AICc	Corrected Akaike information criterion
ANCOVA	Analysis of covariance
CCM	Carbon concentrating mechanism
DOE	Design of experiments
FAMES	Fatty acid methyl esters
IEF	Isoelectrofocalisation
I.S.	Internal standard
LHC	Light-harvesting complex
MLR	Multiple linear regression
mW	Molecular weight
PCA	Principal component analysis
pI	Isoelectric point
PLSR	Partial least squares regression

PRESS	Prediction error sum of squares
$RMSE_F$	Fitting root-mean-squared error
$RMSE_{CV}$	Cross-validation root-mean-squared error
TCA	Tricarboxylic acid
VIP	Variable importance in projection

Conflict of Interest Statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contributions

GM is the author of the original idea of the work. SG and GM conceived the DOE. SG performed algal cultures, proteomic experiments as well as triglyceride and Lichtenthaler's pigment determinations with helpful advice from GM, PL and FES for 2D-DIGE and from FF for spectroscopy. SG and GM carried out chromatographic experiments. SG performed statistical analyses with GM's contribution and wrote the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by a “Fonds de la Recherche Fondamentale et Collective” grant (FRFC 2.4597.11) and a “Fonds de la Recherche Scientifique Médicale” grant (FRSM 3.4559.11) from the Belgian “Fonds de la Recherche Scientifique-Fonds National de la Recherche Scientifique” institution (F.R.S.-FNRS).

Acknowledgements

SG is the recipient of an Aspirant doctoral fellowship from F.R.S.-FNRS. FF is a Research Director and PL is a Senior Research Associate of F.R.S.-FNRS. We thank Professor Edwin De Pauw (Laboratory of Mass Spectrometry, University of Liege) and the "Centre d'Analyse des Résidus en Trace" (CART), GIGA-Research, University of Liege, for protein identifications.

References

Allen, R., Trelease, R., and Thomas, T. (1988). Regulation of isocitrate lyase gene expression in sunflower. *Plant Physiology* 86**,** 527-532.

Badger, M., von Caemmerer, S., Ruuska, S., and Nakano, H. (2000). Electron flow to oxygen in higher plants and algae: rates and control of direct photoreduction (Mehler reaction) and rubisco oxygenase. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* 355**,** 1433-1446.

Bligh, E., and Dyer, W. (1959). A rapid method of total lipid extraction and purification. *Canadian Journal of Biochemistry and Physiology* 37**,** 911-917.

Borkhsenious, O., Mason, C., and Moroney, J. (1998). The intracellular localization of ribulose-1,5-bisphosphate carboxylase/oxygenase in Chlamydomonas reinhardtii. *Plant Physiology* 116**,** 1585-1591.

Boyle, N., and Morgan, J. (2009). Flux balance analysis of primary metabolism in Chlamydomonas reinhardtii. *BMC Systems Biology* 3**,** 4.

Brooks, A., Portis, A., and Sharkey, T. (1988). Effects of irradiance and methyl viologen treatment on ATP, ADP, and activation of ribulose bisphosphate carboxylase in spinach leaves. *Plant Physiology* 88**,** 850-853.

Browse, J., McCourt, P., and Somerville, C. (1986). Fatty acid composition of leaf lipids determined after combined digestion and fatty acid methyl ester formation from fresh tissue. *Analytical Biochemistry* 152**,** 141-145.

Burnham, K., and Anderson, D. (2004). Multimodel inference. Understanding AIC and BIC in model selection. *Sociological Methods and Research* 33**,** 261-304.

Carpentier, S., Witters, E., Laukens, K., Deckers, P., Swennen, R., and Panis, B. (2005). Preparation of protein extracts from recalcitrant plant tissues: an evaluation of different methods for two-dimensional gel electrophoresis analysis. *Proteomics* 5**,** 2497-2507.

Chen, F., and Johns, M. (1994). Substrate inhibition of Chlamydomonas reinhardtii by acetate in heterotrophic culture. *Process Biochemistry* 29**,** 245-252.

Chen, F., and Johns, M. (1996). Heterotrophic growth of Chlamydomonas reinhardtii on acetate in chemostat culture. *Process Biochemistry* 31**,** 601-604.

Collos, Y., and Harrison, P. (2014). Acclimation and toxicity of high ammonium concentrations to unicellular algae. *Marine Pollution Bulletin* 80**,** 8-23.

Durnford, D., Price, J., McKim, S., and Sarchfield, M. (2003). Light-harvesting complex gene expression is controlled by both transcriptional and post-transcriptional mechanisms during photoacclimation in Chlamydomonas reinhardtii. *Physiologia Plantarum* 118**,** 193-205.

Falkowski, P., and LaRoche, J. (1991). Acclimation to spectral irradiance in algae. *Journal of Phycology* 27**,** 8-14.

Falkowski, P., and Raven, J. (2013). *Aquatic photosynthesis.* Princeton, NJ: Princeton University Press.

Fernandez, E., Galvan, A., and Quesada, A. (2004). "Nitrogen assimilation and its regulation," in *The molecular biology of chloroplasts and mitochondria in Chlamydomonas,* ed. Springer-Netherlands.), 637-659.

Finazzi, G., Johnson, G., Dall'Osto, L., Zito, F., Bonente, G., Bassi, R., and Wollman, F. (2006). Nonphotochemical quenching of chlorophyll fluorescence in Chlamydomonas reinhardtii. *Biochemistry* 45**,** 1490-1498.

Foyer, C., and Noctor, G. (2003). Redox sensing and signalling associated with reactive oxygen in chloroplasts, peroxisomes and mitochondria. *Physiologia Plantarum* 119**,** 355-364.

Foyer, C., Noctor, G., and Hodges, M. (2011). Respiration and nitrogen assimilation: targeting mitochondria-associated metabolism as a means to enhance nitrogen use efficiency. *Journal of Experimental Botany* 62**,** 1467-1482.

Fukushima, A., Kusano, M., Redestig, H., Arita, M., and Saito, K. (2009). Integrated omics approaches in plant systems biology. *Current Opinion in Chemical Biology* 13**,** 532-538.

Gérin, S., Mathy, G., Blomme, A., Franck, F., and Sluse, F. (2010). Plasticity of the mitoproteome to nitrogen sources (nitrate and ammonium) in Chlamydomonas

reinhardtii: the logic of Aox1 gene localization. *Biochimica et Biophysica Acta-Bioenergetics* 1797**,** 994-1003.
Gérin, S., Mathy, G., and Franck, F. (2014). Modeling the dependence of respiration and photosynthesis upon light, acetate, carbon dioxide, nitrate and ammonium in Chlamydomonas reinhardtii using design of experiments and multiple regression. *BMC Systems Biology* 8**,** 96.
Graham, I., Denby, K., and Leaver, C. (1994). Carbon catabolite repression regulates glyoxylate cycle gene expression in cucumber. *The Plant Cell* 6**,** 761-772.
Greenbaum, D., Colangelo, C., Williams, K., and Gerstein, M. (2003). Comparing protein abundance and mRNA expression levels on a genomic scale. *Genome Biology* 4**,** 117.
Hahn, D., Kaltenbach, C., and Kück, U. (1998). The Calvin cycle enzyme sedoheptulose-1,7-bisphosphatase is encoded by a light-regulated gene in Chlamydomonas reinhardtii. *Plant Molecular Biology* 36**,** 929-934.
Harris, E. (2001). Chlamydomonas as a model organism. *Annual Review of Plant Physiology and Plant Molecular Biology* 52**,** 363-406.
Hauser, T., Popilka, L., Hartl, F., and Hayer-Hartl, M. 2015. Role of auxiliary proteins in Rubisco biogenesis and function. *Nature Plants* [Online], 1.
Höhner, R., Barth, J., Magneschi, L., Jaeger, D., Niehues, A., Bald, T., Grossman, A., Fufezan, C., and Hippler, M. (2013). The metabolic status drives acclimation of iron deficiency responses in Chlamydomonas reinhardtii as revealed by proteomics based hierarchical clustering and reverse genetics. *Molecular and Cellular Proteomics* 12**,** 2774-2790.
Huppe, H., and Turpin, D. (1994). Integration of carbon and nitrogen metabolism in plant and algal cells. *Annual Review of Plant Physiology and Plant Molecular Biology* 45**,** 577-607.
Hüttemann, M., Lee, I., Samavati, L., Yu, H., and Doan, J. (2007). Regulation of mitochondrial oxidative phosphorylation through cell signaling. *Biochimica et Biophysica Acta-Molecular Cell Research* 1773**,** 1701-1720.
Hyams, J., and Davies, D. (1972). The induction and characterization of cell wall mutants of Chlamydomonas reinhardi. *Mutation Research* 14**,** 381-389.
Jezek, P., Engstova, H., Zackova, M., Vercesi, A., Costa, A., Arruda, P., and Garlid, K. (1998). Fatty acid cycling mechanism and mitochondrial uncoupling proteins. *Biochimica et Biophysica Acta-Bioenergetics* 1365**,** 319-327.
Johnson, X., and Alric, J. (2012). Interactions between starch breakdown, acetate assimilation and cyclic electron flow in Chlamydomonas reinhardtii. *The Journal of Biological Chemistry* 287**,** 26445-26452.
Kupriyanova, E., Sinetova, M., Cho, S., Park, Y., Los, D., and Pronina, N. (2013). CO2-concentrating mechanism in cyanobacterial photosynthesis: organization, physiological role, and evolutionary origin. *Photosynthesis Research* 117**,** 133-146.
Lamb, R., Bonuccelli, G., Ozsvari, B., Peiris-Pages, M., Fiorillo, M., Smith, D., Bevilacqua, G., Mazzanti, C., McDonnell, L., Naccarato, A., Chiu, M., Wynne, L., Martinez-Outschoorn, U., Sotgia, F., and Lisanti, M. (2015). Mitochondrial mass, a new metabolic biomarker for stem-like cancer cells: Understanding WNT/FGF-driven anabolic signaling. *Oncotarget* 6**,** 30453.
Lee, D., Park, J., Barupal, D., and Fiehn, O. (2012). System response of metabolic networks in Chlamydomonas reinhardtii to total available ammonium. *Molecular and Cellular Proteomics* 11**,** 973-988.
Lemaire, S., Guillon, B., Le Maréchal, P., Keryer, E., Miginiac-Maslow, M., and Decottignies, P. (2004). New thioredoxin targets in the unicellular photosynthetic

eurkaryote Chlamydomonas reinhardtii. *Proceeding of the National Academy of Sciences of the USA* 101**,** 7475-7480.
Lichtenthaler, H., and Wellburn, A. (1983). Determinations of total carotenoids and chlorophyll a and b in leaf extracts in different solvents. *Biochemical Society Transactions* 11**,** 591-592.
Lowry, O., Rosebrough, N., Farr, A., and Randall, R. (1951). Protein measurement with the Folin phenol reagent. *Journal of Biological Chemistry* 193**,** 165-175.
Marouga, R., David, S., and Hawkins, E. (2005). The development of the DIGE system: 2D fluorescence difference gel analysis technology. *Analytical and Bioanalytical Chemistry* 382**,** 669-678.
Mathy, G., Cardol, P., Dinant, M., Blomme, A., Gérin, S., Cloes, M., Ghysels, B., DePauw, E., Leprince, P., Remacle, C., Sluse-Goffart, C., Franck, F., Matagne, R., and Sluse, F. (2010). Proteomic and functional characterization of a Chlamydomonas reinhardtii mutant lacking the mitochondrial alternative oxidase 1. *Journal of Proteome Research* 9**,** 2825-2838.
Mathy, G., and Sluse, F. (2008). Mitochondrial comparative proteomics: strengths and pitfalls. *Biochimica et Biophysica Acta-Bioenergetics* 1777**,** 1072-1077.
May, C., Brosseron, F., Chartowski, P., Meyer, H., and Marcus, K. (2012). "Differential proteome analysis using 2D-DIGE," in *Methods in Molecular Biology,* ed. Katrin-Marcus. Humana Press), 75-82.
May, P., Christian, J., Kempa, S., and Walther, D. (2009). ChlamyCyc: an integrative systems biology database and web-portal for Chlamydomonas reinhardtii. *BMC Genomics* 10**,** 209.
Merchant, S., Prochnik, S., Vallon, O., Harris, E., Karpowicz, S., Witman, G., Terry, A., Salamov, A., Fritz-Laylin, L., Maréchal-Drouard, L., Marshall, W., Qu, L.-H., Nelson, D., Sanderfoot, A., Spalding, M., Kapitonov, V., Ren, Q., Ferris, P., Lindquist, E., Shapiro, H., Lucas, S., Grimwood, J., Schmutz, J., Cardol, P., Cerutti, H., Chanfreau, G., Chen, C.-L., Cognat, V., Croft, M., Dent, R., Dutcher, S., Fernandez, E., Fukuzawa, H., Gonzalez-Ballester, D., Gonzalez-Halphen, D., Hallmann, A., and Hanikenne, M. (2007). The Chlamydomonas genome reveals the evolution of key animal and plant functions. *Science* 318**,** 245-250.
Miller, R., Wu, G., Deshpande, R., Vieler, A., Gärtner, K., Li, X., Moellering, E., Zäuner, S., Cornish, A., Liu, B., Bullard, B., Sears, B., Kuo, M.-H., Hegg, E., Shachar-Hill, Y., Shiu, S.-H., and Benning, C. (2010). Changes in transcript abundance in Chlamydomonas reinhardtii following nitrogen deprivation predict diversion of metabolism. *Plant Physiology* 154**,** 1737-1752.
Mitchell, M., Meyer, M., and Griffiths, H. (2014). Dynamics of carbon-concentrating mechanism induction and protein relocalization during the dark-to-light transition in synchronized Chlamydomonas reinhardtii. *Plant Physiology* 166**,** 1073-1082.
Mochida, K., and Shinozaki, K. (2011). Advances in omics and bioinformatics tools for systems analyses of plant functions. *Plant and Cell Physiology* 52**,** 2017-2038.
Moroney, J., Ma, Y., Frey, W., Fusilier, K., Pham, T., Simms, T., DiMario, R., Yang, J., and Mukherjee, B. (2011). The carbonic anhydrase isoforms of Chlamydomonas reinhardtii: intracellular location, expression, and physiological roles. *Photosynthesis Research* 109**,** 133-149.
Murphy, D. (1986). The molecular organization and function of the photosynthetic membranes of higher plants. *Biochimica et Biophysica Acta-Reviews on Biomembranes* 864**,** 33-94.

Neale, P., and Melis, A. (1986). Algal photosynthetic membrane complexes and the photosynthesis-irradiance curve: a comparison of light adaptation responses in Chlamydomonas reinhardtii (Chlorophyta). *Journal of Phycology* 22**,** 531-538.

Nield, J., Redding, K., and Hippler, M. (2004). Remodeling of light-harvesting protein complexes in Chlamydomonas in response to environmental changes. *Eukaryotic Cell* 3**,** 1370-1380.

Nogales, J., Guijo, M., Quesada, A., and Merchan, F. (2004). Functional analysis and regulation of the malate synthase from Chlamydomonas reinhardtii. *Planta* 219**,** 325-331.

Perchorowicz, J., Raynes, D., and Jensen, R. (1981). Light limitation of photosynthesis and cativation of ribulose bisphosphate carboxylase in wheat seedlings. *Proceeding of the National Academy of Sciences of the USA* 78**,** 2985-2989.

Perez-Garcia, O., Escalante, F., de-Bashan, L., and Bashan, Y. (2011). Heterotrophic cultures of microalgae: Metabolism and potential products. *Water Research* 45**,** 11-36.

Petridou, S., Foster, K., and Kindle, K. (1997). Light induces accumulation of isocitrate lyase mRNA in a carotenoid-deficient mutant of Chlamydomonas reinhardtii. *Plant molecular Biology* 33**,** 381-392.

Plumley, F., and Schmidt, G. (1989). Nitrogen-dependent regulation of photosynthetic gene expression. *Proceeding of the National Academy of Sciences of the USA* 86**,** 2678-2682.

Raines, C. (2003). The Calvin cycle revisited. *Photosynthesis Research* 75**,** 1-10.

Rajaram, H., and Apte, S. (2008). Nitrogen status and heat-stress-dependent differential expression of the cpn60 chaperonin gene influences thermotolerance in the cyanobacterium Anabaena. *Microbiology* 154**,** 317-325.

Roberts, M. (2003). 14-3-3 proteins find new partners in plant cell signalling. *Trends in Plant Science* 8**,** 218-223.

Sager, R., and Granick, S. (1953). Nutritional studies with Chlamydomonas reinhardi. *Annals of the New-York Academy of Sciences* 56**,** 831-838.

SAS (2012). *JMP 10 Modeling and Multivariate Methods.* Cary, NC: SAS Institute.

SAS (2013). *JMP 11 Multivariate Methods.* Cary, NC: SAS Institute.

Schroda, M., Vallon, O., Wollman, F., and Beck, C. (1999). A chloroplast-targeted protein 70 (HSP70) contributes to the photoprotection and repair of photosystem II during and after photoinhibition. *The Plant Cell* 11**,** 1165-1178.

Schuster, G., Timberg, R., and Ohad, I. (1988). Turnover of thylakoid photosystem II proteins during photoinhibition of Chlamydomonas reinhardtii. *European Journal of Biochemistry* 177**,** 403-410.

Shevchenko, A., Henrik Tomas, J., Olsen, J., and Mann, M. (1996). In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nature Protocols* 1**,** 2856-2860.

Singh, A., Elvitigala, T., Bhattacharyya-Pakrasi, M., Aurora, R., Ghosh, B., and Pakrasi, H. (2008). Integration of carbon and nitrogen metabolism with energy production is crucial to energy acclimation in the cyanobacterium Synechocystis. *Plant Physiology* 148**,** 467-478.

Smirnoff, N. (2011). Vitamin C: the metabolism and functions of ascorbic acid in plants. *Advances in Botanical Research* 59**,** 107-177.

Spalding, M. (2009). "The CO2-concentrating mechanism and carbon assimilation," in *The Chlamydomonas sourcebook. Organellar and metabolic processes,* ed. D. Stern. 2 ed: Elsevier Academic Press), 257-301.

Spalding, M., Van, K., Wang, Y., and Nakamura, Y. (2002). Acclimation of Chlamydomonas to changing carbon availability. *Functional Plant Biology* 29**,** 221-230.

Sueltemeyer, D., Klug, K., and Fock, H. (1986). Effect of photon fluence rate on oxygen evolution and uptake by Chlamydomonas reinhardtii suspensions grown in ambient and CO2-enriched air. *Plant Physiology* 81**,** 372-375.

Teramoto, H., Nakamori, A., Minagawa, J., and Ono, T. (2002). Light-intensity-dependent expression of Lhc gene family encoding light-harvesting chlorophyll-a/b proteins of photosystem II in Chlamydomonas reinhardtii. *Plant Physiology* 130**,** 325-333.

Tobin, E., and Silverthorne, J. (1985). Light regulation of gene expression in higher plants. *Annual Review of Plant Physiology* 36**,** 569-593.

Trebst, A. (2003). Function of beta-carotene and tocopherol in photosystem II. *Zeitschrift fur Naturforschung C-A Journal of Biosciences* 58**,** 609-620.

Turpin, D. (1991). Effects of inorganic N availability on algal photosynthesis and carbon metabolism. *Journal of Phycology* 27**,** 14-20.

Vance, P., and Spalding, M. (2005). Growth, photosynthesis, and gene expression in Chlamydomonas over a range of CO2 concentrations and CO2/O2 ratios: CO2 regulates multiple acclimation states. *Canadian Journal of Botany* 83**,** 796-809.

Von Gromoff, E., Treier, U., and Beck, C. (1989). Three light-inducible heat shock genes of Chlamydomonas reinhardtii. *Molecular and Cellular Biology* 9**,** 3911-3918.

Wang, Y., Duanmu, D., and Spalding, M. (2011). Carbon dioxide concentrating mechanism in Chlamydomonas reinhardtii: inorganic carbon transport and CO2 recapture. *Photosynthesis Research* 109**,** 115-122.

Ward Jr, J. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American statistical association* 58**,** 236-244.

White, A., and Critchley, C. (1999). Rapid light curves: a new fluorescence method to assess the state of the photosynthetic apparatus. *Photosynthesis Research* 59**,** 63-72.

Work, V., D'Adamo, S., Radakovits, R., Jinkerson, R., and Posewitz, M. (2012). Improving photosynthesis and metabolic networks for the competitive production of phototroph-derived biofuels. *Current Opinion in Biotechnology* 23**,** 290-297.

Xue, X., Gauthier, D., Turpin, D., and Weger, H. (1996). Interactions between photosynthesis and respiration in the green alga Chlamydomonas reinhardtii. Characterization of light-enhanced dark respiration. *Plant Physiology* 112**,** 1005-1014.

Yang, Y., and Gao, K. (2003). Effects of CO2 concentrations on the freshwater microalgae, Chlamydomonas reinhardtii, Chlorella pyrenoidosa and Scenedesmus obliquus (Chlorophyta). *Journal of Applied Phycology* 15**,** 379-389.

Figure legends

Figure 1 Overview of the methodology and results of the present study. PLSR, partial least squares regression. MLR, multiple linear regression. PCA, principal component analysis. ANCOVA, analysis of covariance.

Figure 2 Image of the G-Dye100-labeled internal standard in the Master gel (n°11 in Additional file 1). The spots which were detected by DeCyder 7.0 and which passed the volume restriction filter are encircled. Among them, those that could be identified by mass spectrometry are highlighted in yellow (see also Table 4). The spots that passed the initial PLSR- and MLR-based screening are pointed out by orange arrows with surrounding Master numbers (see also Additional file 3). pI, isoelectric point. MW, molecular weight.

Figure 3 Hierarchical clustering analysis of protein spot abundance pattern upon DOE conditions. Only the spots which passed the initial PLSR- and MLR-based screening were included. Protein abundance is illustrated as a dendrogram with a green-to-red color scale, and the numbering of culture conditions corresponds to that in Additional file 2. A plot illustrating the 2D-distance among the spots is also provided (upper left) to facilitate cluster visualization. The allocation of the additional assays within the different protein clusters was assessed by a separate hierarchical clustering analysis integrating all biological variables. Clust., cluster.

Figure 4 PCA and PLSR analyses of in-cluster regulatory specificities regarding the DOE conditions. PCA was performed with all biological variables in the same analysis whereas one PLSR was performed for each cluster. **(A)** Results of PCA. The score plot was replicated in five copies so as to enable to mark the observations according to the values taken by each environmental variable in DOE. In the loading plot (at the bottom right), vectors of the biological variables are colored according to their respective cluster. **(B)** Biplot-like scheme summarizing (i) the regulatory tendencies observed within each quadrant of the score plot regarding the DOE conditions and (ii) the angular covering by the vectors of each cluster within the loading plot. **(C)** Results of PLSRs as the variable importance in projection (VIP) of the environmental factors for each cluster. The sign of the coefficients within PLSR models is provided for VIP values exceeding 1.

Figure 5 β-weights associated with the statistically significant effects ($p \leq 0.05$) of MLR models for individual biological variables. β-weights are illustrated as a green-to-red color scale; empty cases are for insignificant effects or effects which were not selected by stepwise regression (see Additional file 10 for raw data). Biological variables are classified by cluster (Figure 3); within each cluster, they were sorted such as to facilitate the visual comparison of their respective regulation patterns. Clust., cluster.

Figure 6 Generalized simulation plots for MLR individual modeling of the biological variables. This figure is the key for reading the regulation results summarized in Figure 5. **(A)** Influence profile of the environmental variables according to the type of effect (ordinal, continuous linear or continuous quadratic) in relationship with the sign and magnitude of the associated β-weight(s). **(B)** Second-order interactions between environmental variables (X_1 and X_2) and simulates the incidence of X_1 variation on the influence profile of X_2 in relationship with the value of the β-weight of the interaction. Possible variations of X_2 graph intercept as a function of X_1 are not represented on the schemes.

Figure 7 Metabolic adaptations induced in response to variations of light, carbon and nitrogen in the medium. These schemes represent interpretations deduced from our results, mostly related to changes in protein abundance. **(A)** Influence of nitrate and ammonium concentrations. **(B)** Influence of light intensity and carbon availability (acetate and CO_2). The postulated effects of the environmental factors are colored in blue and pointed out by bold arrows surrounded by specific symbols describing the type of influence: + and - are for linear profiles whereas concave and convex shapes are for quadratic profiles. CETC, chloroplastic electron transport chain; Fd, ferredoxin; G-3-P, glyceraldehyde-3-phosphate; LHC, light-harvesting complex.

Tables

Table 1 Input parameters and selection criteria used for the initial screening of the biological variables.

	PLSR	MLR
	One model per continuous environmental variable	
	Strategy 1	Strategy 2
JMP input (launch panel parameters)	Responses = all biological variables (proteins or assays) Factors = Acetate, Light, NH_4 or NO_3 By: CO_2 Factor search range = 1	Responses = all biological variables (proteins or assays) Factors = Acetate, Light, NH_4 or NO_3 Continuous environmental variable ² Continuous environmental variable *x* CO_2
Number of models	8 models (optimal number of latent factors = 1)	4 models
Selection criterion for the biological variables	≥ 30% of variability explained by the latent factor for at least 1 model over 8 (≥ 19% in the NO_3 models for protein spots)	Statistical significance with $p \leq 0.075$ for at least 1 model over 4 for protein spots / $p \leq 0.05$ for additional assays
	Unique model with all continuous environmental variables	
	Strategy 3	Strategy 4
JMP input (launch panel parameters)	Responses = all biological variables (proteins or assays) Factors = Acetate, Light, NH_4 and NO_3 By: CO_2 Factor search range = 4	Responses = all biological variables (proteins or assays) Factors = Acetate, Light, NH_4 and NO_3 Acetate ², Light ², NH_4 ² and NO_3 ² Acetate *x* CO_2, Light *x* CO_2, NH_4 *x* CO_2 and NO_3 *x* CO_2
Number of models	2 models (optimal number of latent factors = 2 for protein spots; 4 for additional assays)	1 model
Selection criterion for the biological variables	≥ 30% of variability explained by the latent factors in at least 1 model over 2 for protein spots / ≥ 65% for additional assays	Statistical significance of the model with $p \leq 0.1$ for protein spots / $p \leq 0.05$ for additional assays
	Selection of the biological variables encountering the selection criterion for at least 3 strategies over 4	

Table 2 Description of the environmental and biological variables considered in the present work. Design of experiments (DOE) was carried out to determine the combinations of environmental variables for which the corresponding biological variables should be measured (see Additional file 2). For CR, MA_{CYT}, MA_{ALT}, P_{800}, $\varphi PSII_{800}$ and NPQ_{800}, data were collected from a recent publication from our group (Gérin et al., 2014) performed with the same algal strain and experimental conditions, and with a similar DOE.

Environmental variables - Design of experiments			
Variables	**Type**	**Unit**	**Range/Modalities**
Acetate concentration	Continuous	$g.L^{-1}$	0-1
Light intensity	Continuous	$\mu mol_{photons}.m^{-2}.s^{-1}$	0-200
Ammonium concentration	Continuous	mM	0-15
Nitrate concentration	Continuous	mM	0-20
CO_2 concentration	Ordinal	%	0.035 and 1.5

Biological variables - Experimental determination				
Source	**Variables**			**Unit**
2D-DIGE	Cellular abundance of…	All 135 identified protein spots (see Table 4)		Spot volume normalized by the I.S.
GC		Palmitic acid		$\mu g.mg_{proteins}^{-1}$
		Stearic acid		
		Oleic acid		
		γ-linolenic acid		
		Linolenic acid		
Enzymatic assay		Triglycerides		$\mu g.mg_{proteins}^{-1}$
Lichtenthaler's spectroscopic equations		Chlorophyll a		$\mu g.mg_{proteins}^{-1}$
		Chlorophyll b		
		Total carotenoids		
HPLC		Neoxanthin		Peak area.$mg_{proteins}^{-1}$
		Violaxanthin		
		Lutein		
		β-carotene		
Clark's electrode oxymetry	Data from Gérin et al., 2014 for…	CR	(cellular respiration)	$nmolO_2.min^{-1}.mg_{proteins}^{-1}$
		MA_{CYT}	(apparent maximal activity of the cytochrome pathway)	
		MA_{ALT}	(apparent maximal activity of the alternative pathway)	
		P_{800}	(gross photosynthetic O_2 evolution)	
PAM fluorimetry		$\varphi PSII_{800}$	(quantum yield of photosystem II)	Arbitrary
		NPQ_{800}	(non-photochemical quenching of chlorophyll fluorescence)	

Table 3 Design of experiments. The identification number of each item refers to the DOE described in (Gérin et al., 2014), which served as basis to build the present one. The unit of each environmental variable can be found in Table 2.

Identification number	[Acetate]	Light	$[NH_4^+]$	$[NO_3^-]$	$[CO_2]$
1	0	200	0	20	1.5
2	0	200	0	10	1.5
3	0.5	200	0	0	1.5
4	0.5	100	7.5	10	1.5
5	1	200	0	20	0.035
6	1	0	15	0	1.5
8	1	0	0	0	0.035
9	0	0	15	20	1.5
13	0	100	0	0	1.5
15	0	0	0	0	0.035
16	1	200	15	0	1.5
19	1	0	0	20	1.5
20	1	200	15	20	0.035
21	0.5	100	7.5	10	0.035
22	1	0	15	0	0.035
23	0	200	15	20	0.035
24	1	0	0	0	1.5
25	0	200	15	0	0.035
27	1	200	15	0	0.035
29	0	0	7.5	0	1.5
30	0	0	15	20	0.035
31	0	0	15	0	0.035
32	0	200	0	20	0.035
33	0.5	200	15	10	1.5
34	1	0	15	20	0.035
35	1	200	15	20	1.5
36	0	0	0	20	0.035
37	1	200	0	0	0.035
39	0.5	100	7.5	10	0.035
40	0	200	0	0	0.035
41	0.5	100	7.5	10	1.5
42	0.5	100	7.5	10	1.5

Table 4 Results of mass spectrometry identifications. All identified proteins were found to belong to *C. reinhardtii*. Identified spots are sorted by alphabetic order of corresponding Gene Name for visual convenience. The spot volume of the G-Dye100-labeled internal standard (I.S.) in the Master gel is also provided as a reference to assess protein abundance in 2D-gels. The spots which passed the initial PLSR and MLR-based screening are presented in the upper part of the table and highlighted in bold (see also Additional file 3). pI, isoelectric point. MW, molecular weight.

Master number	Gene name	GI number in NCBI	Protein description	pI	MW	I.S. spot volume in Master gel
1	**ACH1**	**gi|159462944**	**aconitate hydratase**	**8.9**	**86754**	**140111**
190	**ACH1**	**gi|159462944**	**aconitate hydratase**	**8.9**	**86754**	**855573**
192	**ACH1**	**gi|159462944**	**aconitate hydratase**	**8.9**	**86754**	**459786**
10	**ACS3**	**gi|159488061**	**acetyl CoA synthetase**	**7.3**	**74089**	**179390**
11	**ACS3**	**gi|159488061**	**acetyl CoA synthetase**	**7.3**	**74089**	**352380**
13	**ACS3**	**gi|159488061**	**acetyl CoA synthetase**	**7.3**	**74089**	**439169**
66	**AGS1**	**gi|159477301**	**argininosuccinate synthase**	**8.4**	**49218**	**780829**
67	**AGS1**	**gi|159477301**	**argininosuccinate synthase**	**8.4**	**49218**	**224729**
220	**ASA1**	**gi|159468466**	**mitochondrial F1F0 ATP synthase associated 60.6 kDa protein**	**5.8**	**63123**	**173474**
86	**AST1**	**gi|159473837**	**aspartate aminotransferase**	**9.7**	**46902**	**149498**
18	**ATP2**	**gi|159466892**	**beta subunit of mitochondrial ATP synthase**	**5.0**	**61954**	**231602**
186	**ATP2**	**gi|159466892**	**beta subunit of mitochondrial ATP synthase**	**5.0**	**61954**	**863821**
38	**ATPA**	**gi|41179050**	**ATP synthase CF1 alpha subunit**	**5.4**	**54832**	**495031**
42	**ATPA**	**gi|41179050**	**ATP synthase CF1 alpha subunit**	**5.4**	**54832**	**1711421**
181	**ATPvE**	**gi|159469570**	**vacuolar ATP synthase subunit E**	**7.5**	**26399**	**230593**
68	**BCR1**	**gi|159488652**	**biotin carboxylase, acetyl-CoA carboxylase component**	**9.0**	**52308**	**494766**
199	**BCR1**	**gi|159488652**	**biotin carboxylase, acetyl-CoA carboxylase component**	**9.0**	**52308**	**493171**
63	**BLD10**	**gi|159489304**	**basal body protein**	**5.0**	**174819**	**3481530**
56	**CAT1**	**gi|159477329**	**catalase/peroxidase**	**6.9**	**56407**	**397278**
57	**CAT1**	**gi|159477329**	**catalase/peroxidase**	**6.9**	**56407**	**107595**
232	**CIS1**	**gi|159490012**	**citrate synthase**	**9.1**	**51376**	**74708**
28	**CPN60A**	**gi|159491478**	**chaperonin 60A**	**5.5**	**61911**	**146530**
29	**CPN60A**	**gi|159491478**	**chaperonin 60A**	**5.5**	**61911**	**309631**
30	**CPN60A**	**gi|159491478**	**chaperonin 60A**	**5.5**	**61911**	**113632**
31	**CPN60A**	**gi|159491478**	**chaperonin 60A**	**5.5**	**61911**	**269507**
126	**CPX1**	**gi|159487437**	**coproporphyrinogen III oxidase**	**9.0**	**41743**	**288842**
108	**CYN38**	**gi|159467709**	**peptidyl-prolyl cis-trans isomerase, cyclophilin-type**	**5.4**	**44781**	**138802**
248	**CYP55B1**	**gi|159484456**	**cytochrome P450, nitric oxide reductase**	**6.5**	**44185**	**210579**
70	**EEF1A1**	**gi|159476938**	**eukaryotic translation elongation factor 1 alpha 1**	**8.7**	**51191**	**2335176**
71	**EEF1A1**	**gi|159476938**	**eukaryotic translation elongation factor 1 alpha 1**	**8.7**	**51191**	**2072864**
72	**EEF1A1**	**gi|159476938**	**eukaryotic translation elongation factor 1 alpha 1**	**8.7**	**51191**	**1071287**
195	**EEF1A1**	**gi|159476938**	**eukaryotic translation elongation factor 1 alpha 1**	**8.7**	**51191**	**587077**
253	**EFTU_III**	**gi|41179007**	**elongation factor Tu**	**5.9**	**45772**	**237513**
116	**FNR1**	**gi|159478523**	**ferredoxin-nadp reductase**	**8.5**	**38698**	**1802268**
182	**FNR1**	**gi|159478523**	**ferredoxin-nadp reductase**	**8.5**	**38698**	**538494**
240	**FNR1**	**gi|159478523**	**ferredoxin-nadp reductase**	**8.5**	**38698**	**452769**
246	**FNR1**	**gi|159478523**	**ferredoxin-nadp reductase**	**8.5**	**38698**	**204137**
17	**FTSH1**	**gi|159465357**	**membrane AAA-metalloprotease**	**5.6**	**77727**	**282826**
147	**FTT2**	**gi|159477028**	**14-3-3 protein**	**4.9**	**28099**	**389454**
193	**GAP3**	**gi|159463282**	**glyceraldehyde-3-phosphate dehydrogenase**	**9.2**	**40507**	**570521**
227	**GAP3**	**gi|159463282**	**glyceraldehyde-3-phosphate dehydrogenase**	**9.2**	**40507**	**557330**
105	**GLN2**	**gi|159469782**	**glutamine synthetase**	**7.1**	**41715**	**310080**
106	**GLN2**	**gi|159469782**	**glutamine synthetase**	**7.1**	**41715**	**155830**
243	**GLN2**	**gi|159469782**	**glutamine synthetase**	**7.1**	**41715**	**335820**
163	**GSTS2**	**gi|159482414**	**glutathione S-transferase**	**5.5**	**23922**	**285290**
15	**HSP70A**	**gi|159486599**	**heat shock protein 70A**	**5.3**	**71513**	**468194**
16	**HSP70A**	**gi|159486599**	**heat shock protein 70A**	**5.3**	**71513**	**611102**
4	**HSP70B**	**gi|159476666**	**heat shock protein 70B**	**5.2**	**72081**	**502751**
6	**HSP70B**	**gi|159476666**	**heat shock protein 70B**	**5.2**	**72081**	**872383**
82	**ICL1**	**gi|159474436**	**isocitrate lyase**	**5.9**	**45948**	**643018**
104	**IDH2**	**gi|159473471**	**isocitrate dehydrogenase - NAD-dependent**	**8.8**	**38796**	**388985**
118	**LHCB5**	**gi|159475641**	**minor chlorophyll a-b binding protein of photosystem II**	**5.4**	**30695**	**608401**
157	**LHCBM1**	**gi|20269804**	**major light-harvesting complex II protein m1**	**6.0**	**27605**	**396217**
135	**LHCBM3**	**gi|159491492**	**light-harvesting complex II chlorophyll a-b binding protein M3**	**5.7**	**27420**	**1168663**
184	**LHCBM3**	**gi|159491492**	**light-harvesting complex II chlorophyll a-b binding protein M3**	**5.7**	**27420**	**3613827**
137	**LHCBM6**	**gi|159474480**	**chloropyll a-b binding protein of LHCII type I, chloroplast precursor**	**5.9**	**27058**	**3345291**
34	**MAS1**	**gi|159475042**	**malate synthase**	**8.7**	**61011**	**126787**
35	**MAS1**	**gi|159475042**	**malate synthase**	**8.7**	**61011**	**149623**
75	**METM**	**gi|159477124**	**S-Adenosylmethionine synthetase**	**6.0**	**43070**	**647025**
77	**MPPA2**	**gi|159465665**	**mitochondrial processing peptidase alpha subunit**	**9.7**	**49559**	**122510**
25	**PCK1a**	**gi|159473685**	**phosphoenolpyruvate carboxykinase - splice variant a**	**6.2**	**62388**	**342592**
27	**PCK1a**	**gi|159473685**	**phosphoenolpyruvate carboxykinase - splice variant a**	**6.2**	**62388**	**568919**
87	**PGK1**	**gi|159482940**	**phosphoglycerate kinase**	**8.9**	**49172**	**913779**
153	**POA1**	**gi|159467074**	**20S proteasome alpha subunit A**	**7.6**	**27487**	**199430**

101	**PRK1**	**gi|159471788**	**phosphoribulokinase**	**9.0**	**42151**	**704307**
172	**PSBP1**	**gi|159471964**	**oxygen-evolving enhancer protein 2 of photosystem II**	**9.8**	**29971**	**6598334**
176	**PSBP1**	**gi|159471964**	**oxygen-evolving enhancer protein 2 of photosystem II**	**9.8**	**29971**	**1332361**
46	**RBCL**	**gi|41179049**	**ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit**	**6.1**	**53193**	**768686**
47	**RBCL**	**gi|41179049**	**ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit**	**6.1**	**53193**	**1962505**
50	**RBCL**	**gi|41179049**	**ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit**	**6.1**	**53193**	**573314**
51	**RBCL**	**gi|41179049**	**ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit**	**6.1**	**53193**	**2761041**
52	**RBCL**	**gi|41179049**	**ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit**	**6.1**	**53193**	**953428**
188	**RBCL**	**gi|41179049**	**ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit**	**6.1**	**53193**	**11015508**
185	**RPSA**	**gi|159489000**	**ribosomal protein Sa, component of cytosolic 80S ribosome and 40S small subunit**	**5.1**	**30971**	**447905**
210	**SEBP1**	**gi|159467635**	**sedoheptulose-1,7-bisphosphatase**	**9.6**	**42393**	**1334623**
249	**SHMT2**	**gi|159487140**	**serine hydroxymethyltransferase 2**	**6.3**	**52228**	**219816**
97	**SNE5**	**gi|159487407**	**NAD-dependent epimerase/dehydratase**	**7.8**	**36568**	**211419**
3	**TRK1**	**gi|159487741**	**transketolase**	**7.1**	**78352**	**1478739**
5	**TRK1**	**gi|159487741**	**transketolase**	**7.1**	**78352**	**216394**
8	**TRK1**	**gi|159487741**	**transketolase**	**7.1**	**78352**	**470044**
112	**UPTG1**	**gi|159471081**	**UDP-Glucose:protein transglucosylase**	**5.9**	**39846**	**91980**
113	**UPTG1**	**gi|159471081**	**UDP-Glucose:protein transglucosylase**	**5.9**	**39846**	**348353**
215	**UPTG1**	**gi|159471081**	**UDP-Glucose:protein transglucosylase**	**5.9**	**39846**	**68875**
191	ACH1	gi|159462944	aconitate hydratase	8.9	86754	783724
231	ACH1	gi|159462944	aconitate hydratase	8.9	86754	164370
12	ACS3	gi|159488061	acetyl CoA synthetase	7.3	74089	106938
80	ASA2	gi|159477287	mitochondrial F1F0 ATP synthase associated 45.5 kDa protein	9.6	48383	131696
103	ASSD1	gi|159473875	aspartate semialdehyde dehydrogenase	9.2	40138	411619
39	ATPA	gi|41179050	ATP synthase CF1 alpha subunit	5.4	54832	743857
40	ATPA	gi|41179050	ATP synthase CF1 alpha subunit	5.4	54832	283203
41	ATPA	gi|41179050	ATP synthase CF1 alpha subunit	5.4	54832	1223927
43	ATPA	gi|41179050	ATP synthase CF1 alpha subunit	5.4	54832	3810462
14	ATPvA1	gi|159480680	vacuolar ATP synthase, subunit A	5.7	68921	132891
59	BCR1	gi|159488652	biotin carboxylase, acetyl-CoA carboxylase component	9.0	52308	354966
60	BLD10	gi|159489304	basal body protein	5.0	174819	1106041
61	BLD10	gi|159489304	basal body protein	5.0	174819	315177
62	BLD10	gi|159489304	basal body protein	5.0	174819	560223
111	CYN38	gi|159467709	peptidyl-prolyl cis-trans isomerase, cyclophilin-type	5.4	44781	528179
79	EFTU_III	gi|41179007	elongation factor Tu	5.9	45772	751191
100	FBA3	gi|159485250	fructose-1,6-bisphosphate aldolase	8.9	41301	2007141
194	FBA3	gi|159485250	fructose-1,6-bisphosphate aldolase	8.9	41301	1881225
200	FBA3	gi|159485250	fructose-1,6-bisphosphate aldolase	8.9	41301	357707
219	FBP1	gi|159465323	fructose-1,6-bisphosphatase	5.6	44929	203216
202	FTSH1	gi|159465357	membrane AAA-metalloprotease	5.6	77727	119047
23	FTSH2	gi|159478022	membrane AAA-metalloprotease	6.2	74509	376353
170	GAD1	gi|159491066	UDP-D-glucuronic acid decarboxylase	8.7	37274	259532
152	GBP1	gi|159463672	G-strand telomere binding protein 1	7.6	24160	702757
58	GCSL	gi|159474092	dihydrolipoyl dehydrogenase	9.3	52905	175803
93	IDA5	gi|159482014	actin	5.3	42094	251251
78	IF4A	gi|159466510	eukaryotic initiation factor 4A-like protein	5.5	47309	136610
129	IPY1	gi|159489184	inorganic pyrophosphatase	6.4	31342	1113052
122	LHCB5	gi|159475641	minor chlorophyll a-b binding protein of photosystem II	5.4	30695	1712493
205	LHCBM1	gi|20269804	major light-harvesting complex II protein m1	6.0	27605	2194002
119	MDH1	gi|159469941	malate dehydrogenase	8.5	36864	1282008
120	MDH1	gi|159469941	malate dehydrogenase	8.5	36864	262435
37	MMSDH	gi|159475673	methylmalonate semi-aldehyde dehydrogenase	8.1	58580	182674
146	PDI2	gi|159462776	protein disulfide isomerase	8.8	27447	176123
88	PGK1	gi|159482940	phosphoglycerate kinase	8.9	49172	1979826
221	PGM1b	gi|159476226	phosphoglycerate mutase	5.6	60921	161753
102	PRK1	gi|159471788	phosphoribulokinase	9.0	42151	323308
216	PRK1	gi|159471788	phosphoribulokinase	9.0	42151	173323
183	PSBO	gi|159473144	oxygen-evolving enhancer protein 1 of photosystem II	8.3	30732	5260813
33	PYK1	gi|159469714	pyruvate kinase	6.7	55233	216722
64	QCR1	gi|159477849	ubiquinol:cytochrome c oxidoreductase 50 kDa core 1 subunit	5.9	55248	115110
48	RBCL	gi|41179049	ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit	6.1	53193	163996
187	RBCL	gi|41179049	ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit	6.1	53193	3590194
73	THS1	gi|159480894	threonine synthase	9.4	54835	126623
7	TRK1	gi|159487741	transketolase	7.1	78352	460717
55	TUA1	gi|159467393	alpha tubulin 1	5.0	50182	283438
45	TUB2	gi|159471706	beta tubulin 2	4.7	50157	341849
53	TUB2	gi|159471706	beta tubulin 2	4.7	50157	514491
54	TUB2	gi|159471706	beta tubulin 2	4.8	50157	478728
74	*nd*	*gi|159468534*	*predicted protein*	*6.3*	*42690*	*575642*
107	*nd*	*gi|159478405*	*hypothetical protein CHLREDRAFT_185022*	*5.5*	*36849*	*115745*
224	*nd*	*gi|159491024*	*hypothetical protein*	*10.3*	*33272*	*94742*

Table 5 Protein function(s) and sub-cellular localization(s) as found in the ChlamyCyc database. Proteins are denominated by their corresponding Gene Name (see Table 4 for a complete description) and classified by cluster for visual convenience.

Gene name	Function(s)	Location(s)
	Cluster 1	
RBCL, TRK1	Calvin cycle	chloroplast
	Cluster2	
BCR1	fatty acid biosynthesis	chloroplast
CYP55B1	nitric oxide detoxification	*nd*
EFTU_III	protein elongation	chloroplast
FTSH1	photosystem maintenance	chloroplast
ICL1	glyoxylate cycle	mitochondrion, peroxisome
LHCB5	light-harvesting antennae	chloroplast
POA1	proteasome	cytosol
PSBP1	photosynthetic O_2 evolution	chloroplast
	Cluster3	
CPN60A	protein folding and stability	chloroplast
	Cluster4	
ACH1	TCA cycle, glyoxylate cycle	mitochondrion
ACS3	acetate conversion to acetyl-CoA	cytosol, mitochondrion
AST1	amino-acid interconversion, anaplerosis, malate-oxaloacetate shuttle	chloroplast, mitochondrion
CAT1	H_2O_2 detoxification	mitochondrion, peroxisome
MAS1	glyoxylate cycle	peroxisome
PCK1a	gluconeogenesis	cytosol
PGK1	glycolysis, gluconeogenesis, Calvin cycle	chloroplast
	Cluster 5	
ATPA	ATP synthase, F_1 subunit component	chloroplast, thylakoid membrane
BLD10	flagellum assembly and structure	cytosol
GLN2	GS/GOGAT cycle	chloroplast
HSP70A	protein folding and stabilization	cytosol
HSP70B	photosystem assembly and maintenance	chloroplast
METM	S-adenosylmethionine biosynthesis	cytosol, mitochondrion
MPPA2	protein import to mitochondria	mitochondrion
SNE5	cell-wall and secondary metabolite biosynthesis	*nd*
UPTG1	protein glycosylation	cytosol, mitochondrion
	Cluster 6	
BCR1	fatty acid biosynthesis	chloroplast
EEF1A1, RPSA	protein elongation	cytosol
GAP3, PRK1, SEBP1	Calvin cycle	chloroplast
SHMT2	photorespiration	mitochondrion, cytosol
	Cluster 7	
AGS1	arginine biosynthesis	chloroplast
ASA1	ATP synthase, F_1 subunit component	mitochondrion
ATP2	ATP synthase, F_1 subunit component	mitochondrion, inner membrane
ATPA	ATP synthase, F_1 subunit component	chloroplast, thylakoid membrane
ATPvE	ATP-dependent proton pump for active transport processes	vacuolar membrane
CIS1, IDH2	TCA cycle, glyoxylate cycle	mitochondrion
CPX1	chlorophyll and heme biosynthesis	chloroplast
FNR1	photosynthetic electron transport	chloroplast
FTT2	enzymatic activity regulation	mitochondrion
GSTS2	peroxidized lipids and proteins detoxification	*nd*
	Cluster 8	
CYN38	photosystem assembly and stabilization	chloroplast stroma, thylakoid
FNR1	photosynthetic electron transport	chloroplast
LHCBM1, LHCBM3, LHCBM6	light-harvesting antennae	chloroplast, thylakoid membrane

1

Figure 1

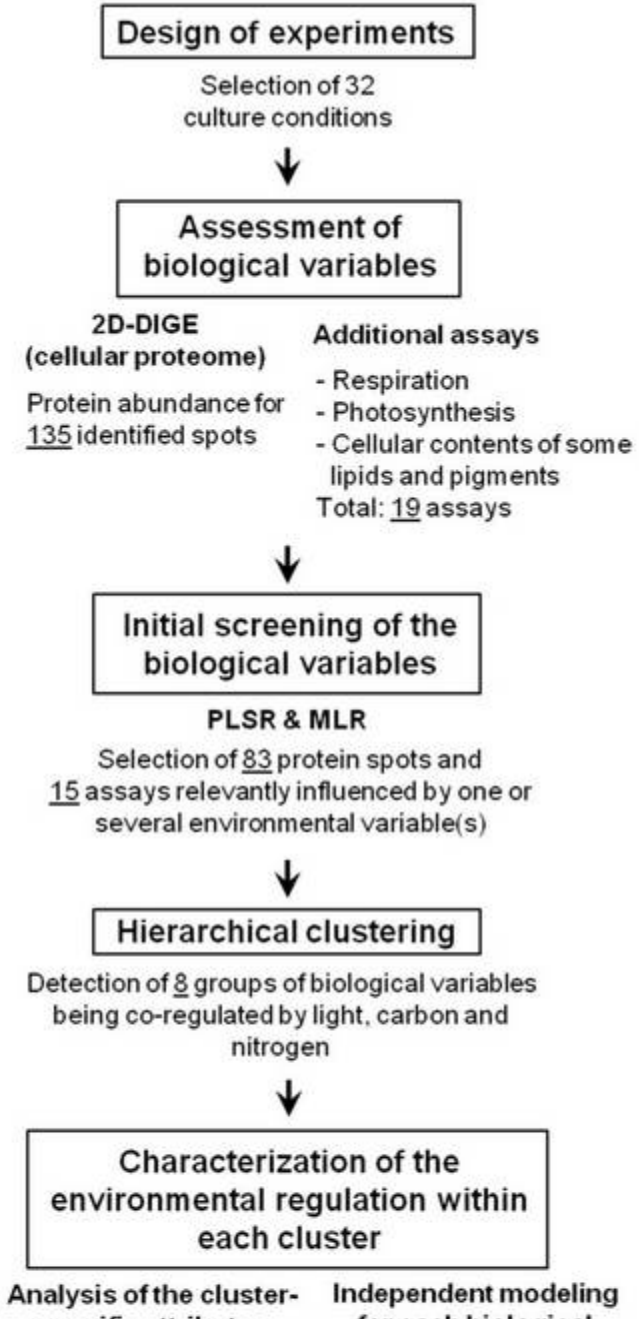

Design of experiments
Selection of 32 culture conditions
Assessment of biological variables
2D-DIGE (cellular proteome)
Protein abundance for 135 identified spots
Additional assays
- Respiration
- Photosynthesis
- Cellular contents of some lipids and pigments
Total: 19 assays
Initial screening of the biological variables
PLSR & MLR
Selection of 83 protein spots and 15 assays relevantly influenced by one or several environmental variable(s)
Hierarchical clustering
Detection of 8 groups of biological variables being co-regulated by light, carbon and nitrogen
Characterization of the environmental regulation within each cluster
Analysis of the cluster-
Independent modeling

Figure 2

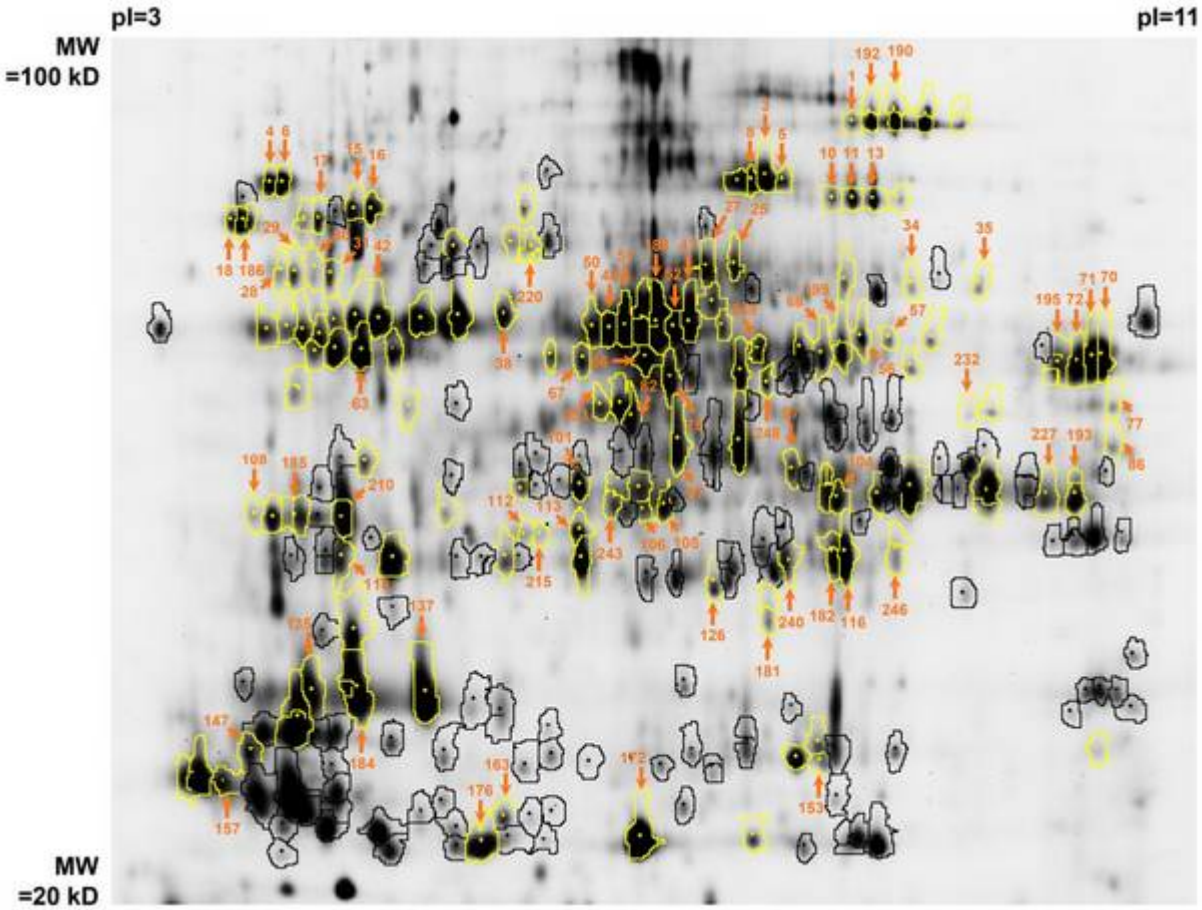


Figure 3

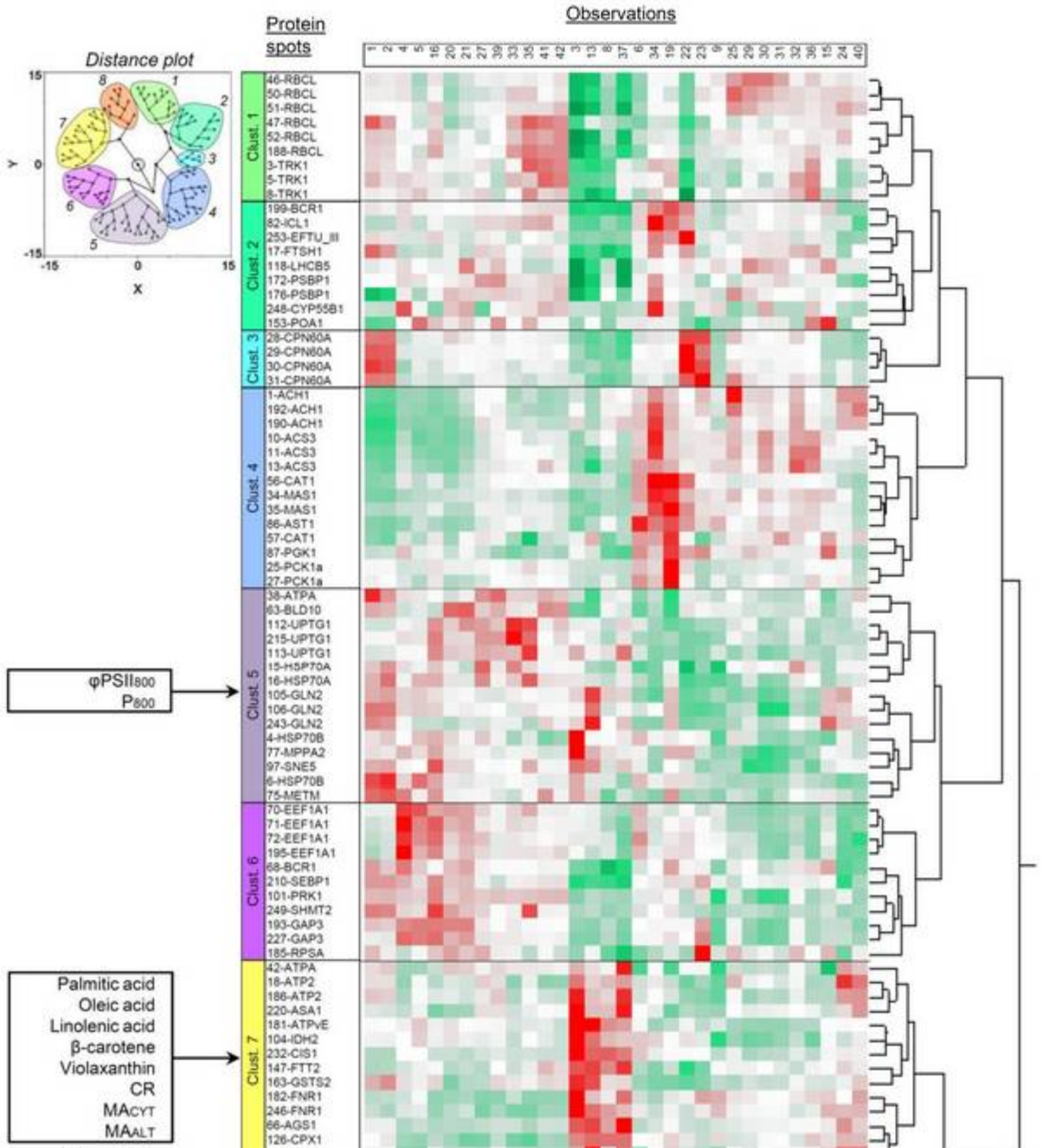

Distance plot
Protein spots
Observations
Clust. 1
46-RBCL
50-RBCL
51-RBCL
47-RBCL
52-RBCL
188-RBCL
3-TRK1
5-TRK1
8-TRK1
Clust. 2
199-BCR1
82-ICL1
253-EFTU_III
17-FTSH1
118-LHCB5
172-PSBP1
176-PSBP1
248-CYP55B1
153-POA1
Clust. 3
28-CPN60A
29-CPN60A
30-CPN60A
31-CPN60A
Clust. 4
1-ACH1
192-ACH1
190-ACH1
10-ACS3
11-ACS3
13-ACS3
56-CAT1
34-MAS1
35-MAS1
86-AST1
57-CAT1
87-PGK1
25-PCK1a
27-PCK1a
Clust. 5
38-ATPA
63-BLD10
112-UPTG1
215-UPTG1
113-UPTG1
15-HSP70A
16-HSP70A
105-GLN2
106-GLN2
243-GLN2
4-HSP70B
77-MPPA2
97-SNE5
6-HSP70B
75-METM
Clust. 6
70-EEF1A1
71-EEF1A1
72-EEF1A1
195-EEF1A1
68-BCR1
210-SEBP1
101-PRK1
249-SHMT2
193-GAP3
227-GAP3
185-RPSA
Clust. 7
42-ATPA
18-ATP2
186-ATP2
220-ASA1
181-ATPvE
104-IDH2
232-CIS1
147-FTT2
163-GSTS2
182-FNR1
246-FNR1
66-AGS1
126-CPX1
φPSII800
P800
Palmitic acid
Oleic acid
Linolenic acid
β-carotene
Violaxanthin
CR
MACYT
MAALT

Figure 4

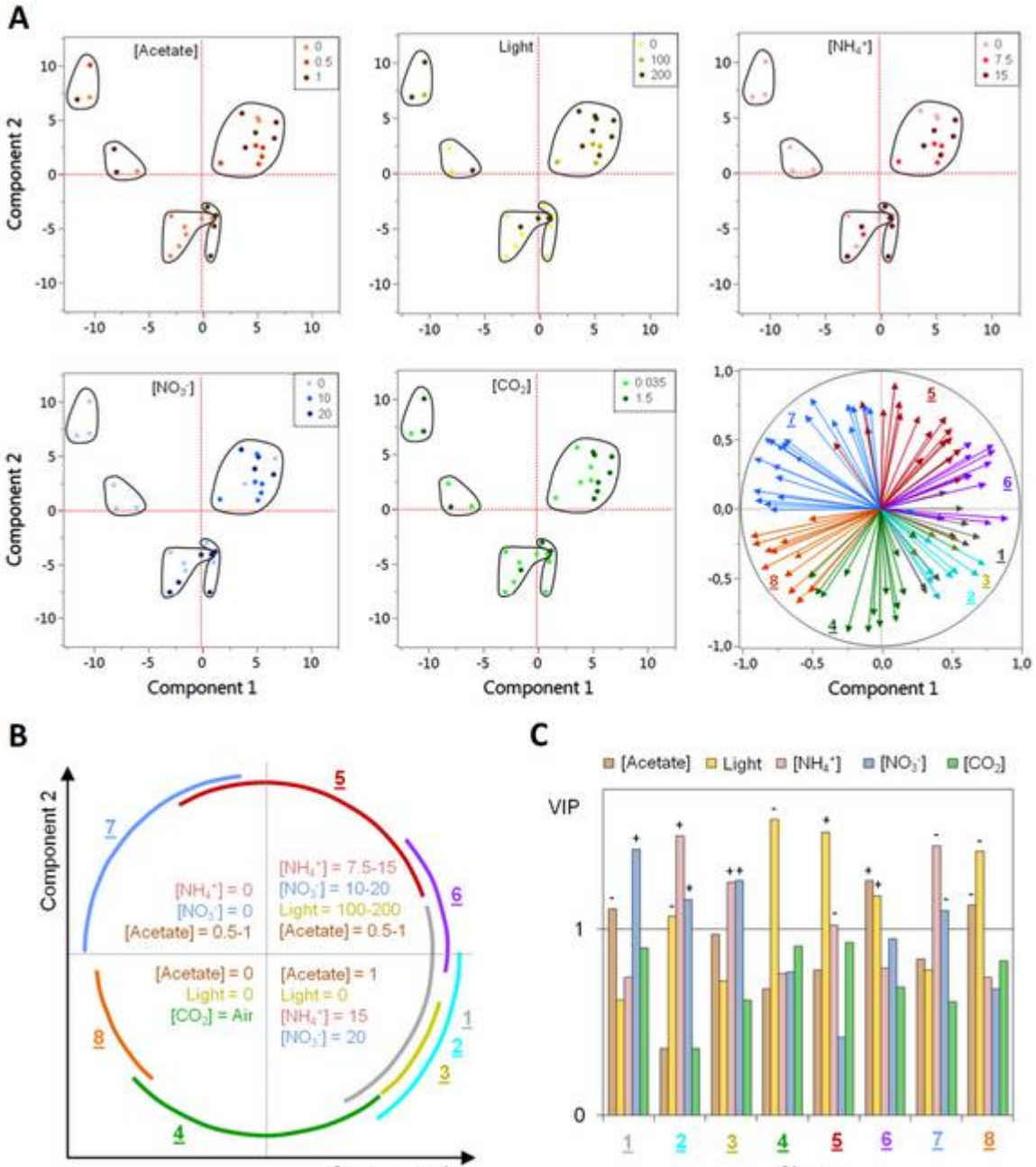

A
[Acetate]
Light
[NH4+]
[NO3-]
[CO2]
Component 2
Component 1
B
[NH4+] = 0
[NO3-] = 0
[Acetate] = 0.5-1
[NH4+] = 7.5-15
[NO3-] = 10-20
Light = 100-200
[Acetate] = 0.5-1
[Acetate] = 0
Light = 0
[CO2] = Air
[Acetate] = 1
Light = 0
[NH4+] = 15
[NO3-] = 20
C
[Acetate] Light [NH4+] [NO3-] [CO2]
VIP

Figure 5

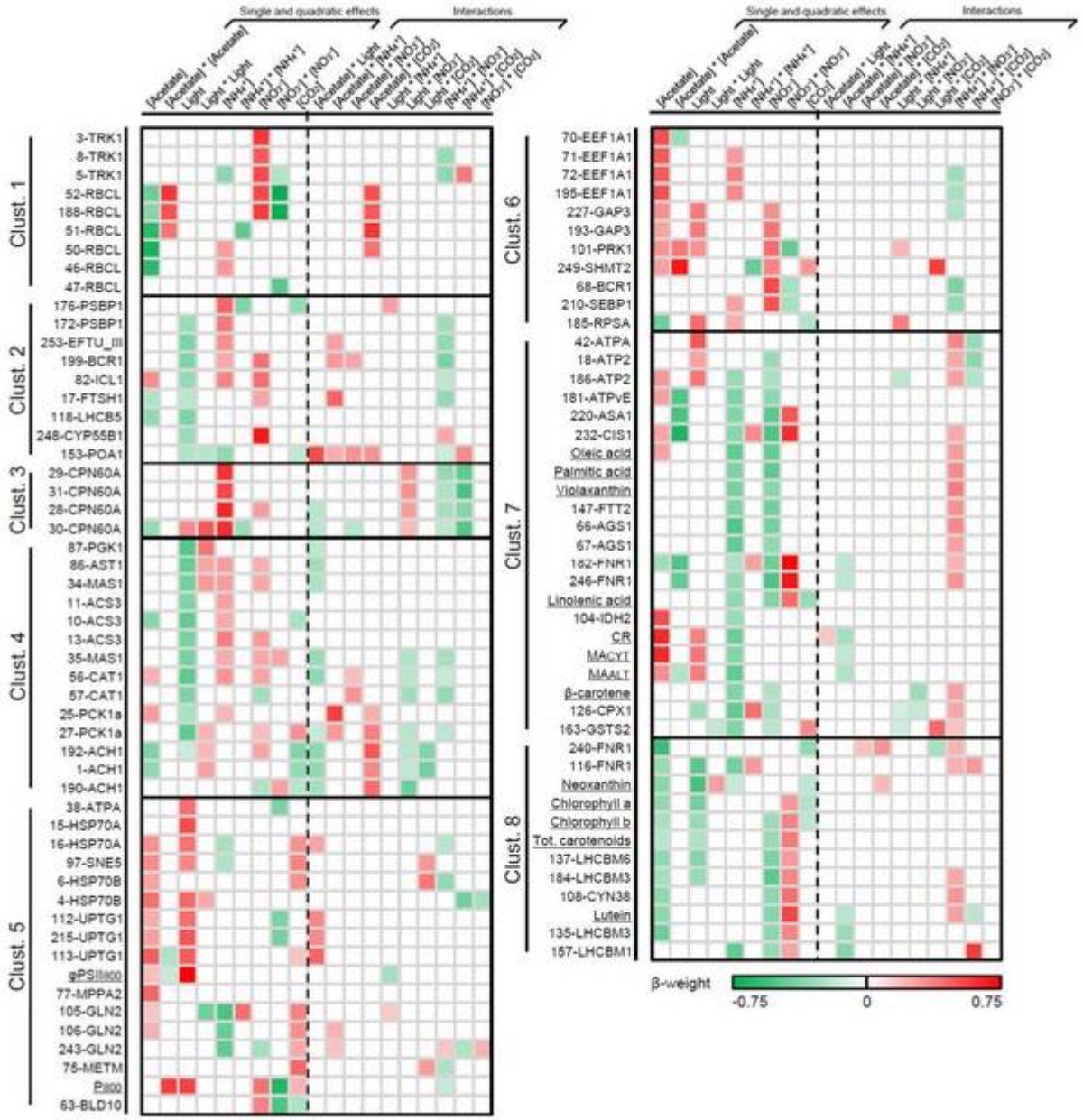

Single and quadratic effects
Interactions
[Acetate]
[Acetate]²
Light
Light²
[NH4+]
[NH4+]²
[NO3-]
[NO3-]²
[CO2]
[Acetate] * Light
[Acetate] * [NH4+]
[Acetate] * [NO3-]
[Acetate] * [CO2]
Light * [NH4+]
Light * [NO3-]
Light * [CO2]
[NH4+] * [NO3-]
[NH4+] * [CO2]
[NO3-] * [CO2]
Clust. 1
3-TRK1
8-TRK1
5-TRK1
52-RBCL
188-RBCL
51-RBCL
50-RBCL
46-RBCL
47-RBCL
Clust. 2
176-PSBP1
172-PSBP1
253-EFTU_III
199-BCR1
82-ICL1
17-FTSH1
118-LHCB5
248-CYP55B1
153-POA1
Clust. 3
29-CPN60A
31-CPN60A
28-CPN60A
30-CPN60A
Clust. 4
87-PGK1
86-AST1
34-MAS1
11-ACS3
10-ACS3
13-ACS3
35-MAS1
56-CAT1
57-CAT1
25-PCK1a
27-PCK1a
192-ACH1
1-ACH1
190-ACH1
Clust. 5
38-ATPA
15-HSP70A
16-HSP70A
97-SNE5
6-HSP70B
4-HSP70B
112-UPTG1
215-UPTG1
113-UPTG1
φPSII800
77-MPPA2
105-GLN2
106-GLN2
243-GLN2
75-METM
P800
63-BLD10
Clust. 6
70-EEF1A1
71-EEF1A1
72-EEF1A1
195-EEF1A1
227-GAP3
193-GAP3
101-PRK1
249-SHMT2
68-BCR1
210-SEBP1
185-RPSA
Clust. 7
42-ATPA
18-ATP2
186-ATP2
181-ATPvE
220-ASA1
232-CIS1
Oleic acid
Palmitic acid
Violaxanthin
147-FTT2
66-AGS1
67-AGS1
162-FNR1
246-FNR1
Linolenic acid
104-IDH2
CR
MACYT
MAALT
β-carotene
126-CPX1
163-GSTS2
Clust. 8
240-FNR1
116-FNR1
Neoxanthin
Chlorophyll a
Chlorophyll b
Tot. carotenoids
137-LHCBM6
184-LHCBM3
108-CYN38
Lutein
135-LHCBM3
157-LHCBM1
β-weight
-0.75
0
0.75

Figure 6

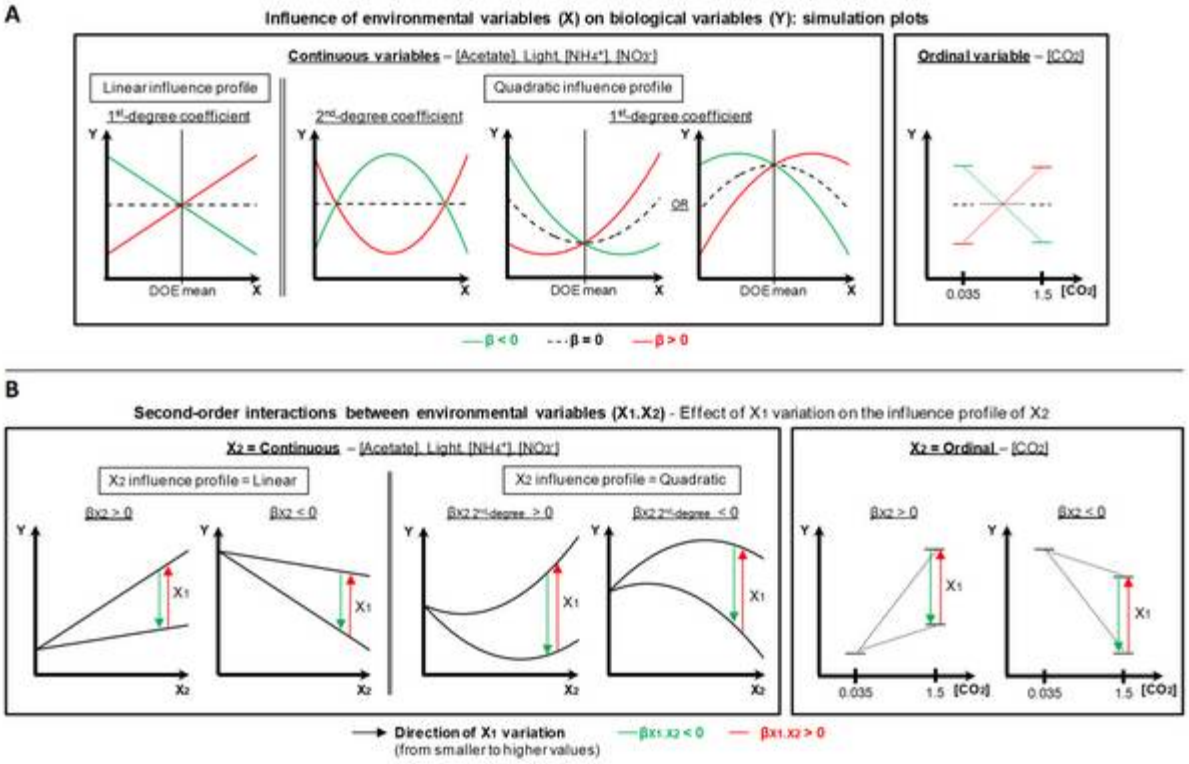
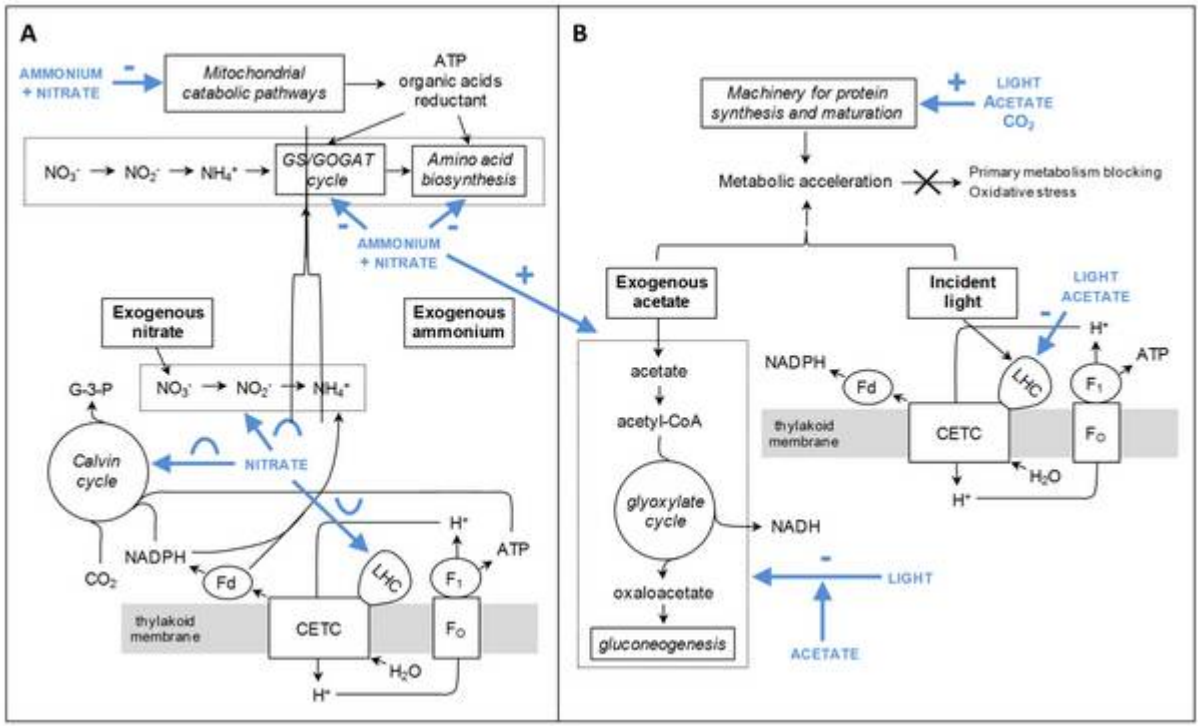

A
Influence of environmental variables (X) on biological variables (Y): simulation plots
Continuous variables – [Acetate], Light, [NH4+], [NOx-]
Linear influence profile
1st-degree coefficient
Quadratic influence profile
2nd-degree coefficient
1st-degree coefficient
OR
DOE mean
Ordinal variable – [CO2]
0.035
1.5 [CO2]
β < 0
β = 0
β > 0
B
Second-order interactions between environmental variables (X1.X2) - Effect of X1 variation on the influence profile of X2
X2 = Continuous – [Acetate], Light, [NH4+], [NOx-]
X2 influence profile = Linear
X2 influence profile = Quadratic
X2 = Ordinal – [CO2]
Direction of X1 variation
(from smaller to higher values)

Figure 7


A
AMMONIUM + NITRATE
Mitochondrial catabolic pathways
ATP
organic acids
reductant
NO3- → NO2- → NH4+
GS/GOGAT cycle
Amino acid biosynthesis
AMMONIUM + NITRATE
Exogenous nitrate
Exogenous ammonium
NO3- → NO2- → NH4+
G-3-P
Calvin cycle
NITRATE
NADPH
Fd
CO2
H+
ATP
LHC
F1
thylakoid membrane
CETC
FO
H2O
H+
B
Machinery for protein synthesis and maturation
LIGHT
ACETATE
CO2
Metabolic acceleration
Primary metabolism blocking
Oxidative stress
Exogenous acetate
Incident light
LIGHT
ACETATE
acetate
acetyl-CoA
glyoxylate cycle
NADH
oxaloacetate
gluconeogenesis
LIGHT
ACETATE
NADPH
Fd
LHC
F1
ATP
H+
thylakoid membrane
CETC
FO
H2O
H+