



Published in final edited form as:

Plant J. 2015 December ; 84(6): 1239–1256. doi:10.1111/tpj.13059.

A refined genome-scale reconstruction of *Chlamydomonas* metabolism provides a platform for systems-level analyses

Saheed Imam¹, Sascha Schäuble^{1,2,3}, Jacob Valenzuela¹, Adrián López García de Lomana¹, Warren Carter¹, Nathan D. Price^{1,4,6}, and Nitin S. Baliga^{1,5,6,7,*}

¹Institute for Systems Biology, Seattle, WA, USA

²Jena University Language & Information Engineering (JULIE) Lab, Friedrich-Schiller-University Jena, Jena, Germany

³Research Group Theoretical Systems Biology, Friedrich-Schiller-University Jena, 07743 Jena, Germany

⁴Departments of Bioengineering and Computer Science & Engineering, University of Washington, Seattle, WA, USA

⁵Departments of Biology and Microbiology, University of Washington, Seattle, WA, USA

⁶Molecular and Cellular Biology Program, University of Washington, Seattle, WA, USA

⁷Lawrence Berkeley National Lab, Berkeley, CA

Summary

Microalgae have reemerged as organisms of prime biotechnological interest due to their ability to synthesize a suite of valuable chemicals. To harness the capabilities of these organisms, we need a comprehensive systems-level understanding of their metabolism, which can be fundamentally achieved through large-scale mechanistic models of metabolism. In this study, we present a revised and significantly improved genome-scale metabolic model for the widely-studied microalga, *Chlamydomonas reinhardtii*. The model, *iCre1355*, represents a major advance over previous models, both in content and predictive power. *iCre1355* encompasses a broad range of metabolic functions encoded across the nuclear, chloroplast and mitochondrial genomes accounting for 1355 genes (1460 transcripts), 2394 and 1133 metabolites. We found improved performance over the previous metabolic model based on comparisons of predictive accuracy across 306 phenotypes (from 81 mutants), lipid yield analysis and growth rates derived from chemostat-grown cells (under 3 conditions). Measurement of macronutrient uptake revealed carbon and phosphate to be good predictors of growth rate, while nitrogen consumption appeared to be in excess. We analyzed high-resolution time series transcriptomics data using *iCre1355* to uncover dynamic pathway-level changes that occur in response to nitrogen starvation and changes in light intensity. This approach enabled accurate prediction of growth rates, the cessation of growth and accumulation of triacylglycerols during nitrogen starvation, and the temporal response of different growth-associated pathways to increased light intensity. Thus, *iCre1355* represents an

*Correspondence: Nitin S. Baliga, Institute for Systems Biology, 401 Terry Ave N., Seattle, WA 98109, Telephone: 206.732.1266, Fax: 206.732.1299, nbaliga@systemsbiology.org.

experimentally-validated genome-scale reconstruction of *C. reinhardtii* metabolism that should serve as a useful resource for studying the metabolic processes of this and related microalgae.

Keywords

Metabolic modeling; *Chlamydomonas reinhardtii*; constraint-based analysis; flux balance analysis; data integration; systems biology; biomass; lipid accumulation; photosynthesis

Introduction

Human activities are continually reshaping the global ecosystem. Of particular concern is the continually increasing level of atmospheric and dissolved CO₂, arising in part, from our dependence of fossil fuels for power (Lewis and Nocera, 2006; Sayre, 2010). The predicted consequences of these activities such as global warming, climate change and ocean acidification, could be detrimental to both terrestrial and aquatic habitats (Feely et al., 2004; Lewis and Nocera, 2006; Sayre, 2010; Honisch et al., 2012; McGlade and Ekins, 2015). Due to their ability to harness solar energy and fix CO₂, photosynthetic organisms are key to future goals of building sustainable societies. Photosynthetic microbes, and microalgae in particular, hold great potential as bio-factories for the economical production of renewable biofuels and a wide variety of other valuable commodities because of their relatively high photosynthetic efficiency, capacity for CO₂ capture (Spalding, 2008; Sayre, 2010), and rapid growth rate. However, the yields of biofuel precursors obtained from microalgae are currently too low to be an economically viable replacement for fossil fuels (Scott et al., 2010; Wijffels and Barbosa, 2010; Reijnders et al., 2014). To overcome this hurdle, an in-depth understanding of microalgae biology including the key metabolic and regulatory components that control relevant processes is required to rationally engineer strains with improved characteristics. We are employing systems-level approaches to gain a deeper understanding of metabolic and regulatory processes in microalgae, with the aim of rationally engineering and enhancing production of value-added commodities.

Chlamydomonas reinhardtii is arguably the best characterized microalga, and has served as a model organism for the study of many cellular and plant-specific processes including photosynthesis, motility, circadian rhythm and the cell cycle control (Harris, 2001; Rochaix, 2002; Marshall, 2008). *C. reinhardtii* has also been used to study the neutral lipid accumulation trait displayed by many microalgae when subjected to abiotic stresses such as nutrient limitation (Boyle et al., 2012; Urzica et al., 2013). The *C. reinhardtii* genome, which was first sequenced in 2005, is the best annotated and curated of any algal species (Merchant et al., 2007; Blaby et al., 2014), placing it at the forefront of algal genomics research. These considerations, along with the availability of tools for genetic manipulation and a large and growing collection of genome-wide datasets (Miller et al., 2010; Boyle et al., 2012; Blaby et al., 2013; Duanmu et al., 2013; Hemschemeier et al., 2013; Mettler et al., 2014; Park et al., 2015), make *C. reinhardtii* the ideal system for studying algal metabolism.

Constraint-based metabolic modeling, which bypasses the need for kinetic parameters that are typically unavailable, provides a useful option for modeling metabolic networks at the systems level (Varma and Palsson, 1994; Palsson, 2000). These models have proven useful

for the design of strains with improved capacity to produce targeted metabolites (Alper et al., 2005; Park et al., 2007; Milne et al., 2009). Furthermore, these models can serve as versatile platforms for integration and contextualization of high-throughput datasets, which can result in new biological insights (Shlomi et al., 2008; Colijn et al., 2009; Oberhardt et al., 2009; Kim and Reed, 2012; McCloskey et al., 2013). Consequently, models of core metabolism (Boyle and Morgan, 2009; Kliphuis et al., 2011) and a genome-scale model (Chang et al., 2011) have previously been constructed for *C. reinhardtii* and used to study different aspects of microalgal metabolism. However, significant improvements in annotation resulting from re-sequencing of the *C. reinhardtii* genome combined with better gene models (Blaby et al., 2014), provide ample data to refine and extend previous *C. reinhardtii* metabolic models. Furthermore, a wide variety of new experimental data providing genetic characterization of biosynthetic pathways (Lin et al., 2010; Leclerc et al., 2012; Urzica et al., 2012; Duanmu et al., 2013; Moulin et al., 2013), as well as a growing list of high-throughput genome-wide measurements from a diverse set of nutritional, environmental and genetic backgrounds (Bolling and Fiehn, 2005; Boyle et al., 2012; Msanne et al., 2012; Blaby et al., 2013; Duanmu et al., 2013; Hemschemeier et al., 2013; Mettler et al., 2014; Park et al., 2015), provide abundant information for model refinement, data integration and systems-level analysis of important biological processes.

In this study, we present a revised and significantly improved genome-scale model for *C. reinhardtii*, *iCre1355*, which represents a major advance over previous models both in content and predictive power. *iCre1355* accounts for a large amount of new and previously omitted information about *C. reinhardtii*'s metabolic capabilities obtained from both genomic and literature sources. We also show that *iCre1355* is a better predictor of gene essentiality, growth rate and lipid yield than the previous genome-scale model. By integrating data from high-throughput transcriptomics experiments, we used *iCre1355* to gain insights into key processes including the metabolic response to nitrogen starvation and the metabolic adjustments associated with growth under different light regimes. Thus, *iCre1355* represents a high-quality, experimentally validated genome-scale metabolic model and a powerful resource for performing systems-level analysis of *C. reinhardtii* and related microalgae.

Results and Discussion

Reconstructing the *C. reinhardtii* metabolic network leveraging new genomic and metabolic information

Updating the *C. reinhardtii* metabolic parts list—Prior to generating constraint-based models for simulating metabolism, the components of the metabolic network (reactions, metabolites, proteins and genes) for which evidence exists need to be systematically compiled into a metabolic reconstruction (Thiele and Palsson, 2010). Using the previously published genome-scale reconstruction for *C. reinhardtii* (*iRC1080*) (Chang et al., 2011) as a reference, we integrated *C. reinhardtii*-specific metabolic and genomic information from 4 databases: the Plant metabolic network (ChlamyCyc 4.0), KEGG (Kanehisa et al., 2002; Kanehisa et al., 2014), NCBI and Phytozome 10.1, to identify a total of 586 candidate metabolic genes to be considered for integration into a refined genome-

scale model (Figure S1). These candidate metabolic genes were then manually curated for literature evidence supporting their inclusion in the reconstruction and/or the presence of appropriate functional domains to carry out associated reactions (see Experimental procedures for reconstruction details). This process led to the pruning of the initial list of 586 candidates to a set of 312 genes with literature- or sequence-based evidence for their inclusion in the model (Table S1, Table S2).

The subsystem distribution of these 312 genes shows that they cover a wide array of key metabolic functions, ranging from central metabolism and photosynthesis to amino acid and cofactor biosynthesis (Figure 1a). Of particular note is the fact that previous *C. reinhardtii* metabolic reconstructions did not consider the biosynthetic pathways relevant for the formation of several essential cofactors including NAD(H), NADP(H), FAD, biotin and thiamine. These pathways were either incomplete or completely missing from other reconstructions, despite the fact that several of them are well characterized in *C. reinhardtii* (Croft et al., 2006; Lin et al., 2010; Moulin et al., 2013). The cofactor biosynthetic pathways were carefully curated and fully accounted for in the current reconstruction. Furthermore, metabolic genes encoded in the chloroplast and mitochondrial genomes, many of which code for essential functions, were also not considered in previous reconstructions. These are now captured in the current reconstruction (Table S1). Given that almost all chloroplast and mitochondrial genes have been genetically and phenotypically characterized, this information was used to establish the gene-protein-reaction (GPR) relationships to nuclear encoded genes involved in the same processes. Overall, these 312 genes are associated with 551 reactions distributed across 8 compartments (Figure 1b).

Refining components from iRC1080—In addition to curating a large number of new metabolic genes and their associated reactions for inclusion in the updated model, we thoroughly re-evaluated the components included in the previously published genome-scale mode for *C. reinhardtii*, iRC1080. Firstly, the gene and transcript identifiers from version 4.3.1 of the *C. reinhardtii* genome used in iRC1080 were updated to the new identifiers used in the current version of the genome (version 5.5 at the time of writing) (Table S3). This process led to the identification of 12 genes and 24 transcripts from iRC1080 that are now obsolete, as they no longer map to any genes (or transcripts) in the current annotated genome (Table S4). We then conducted a comprehensive functional domain analysis of the 1062 genes included in iRC1080, which could be mapped to loci in the new genome annotation. This analysis resulted in the identification of 19 genes that had been incorrectly assigned to various functions in iRC1080 (Table S4). These genes and their associated reactions (in instances where alternative correctly assigned isozymes did not exist) were excluded from further consideration (Table S4).

Modifications were also made to several other reactions in iRC1080. In particular, several erroneous GPR assignments were corrected (Table S4). For instance, the GPR association of the ribulose-1,5-bisphosphate carboxylase oxidase (RuBisCO) reaction in iRC1080 involved an ‘AND’ relationship between the two genes encoding isozymes of the small subunit of RuBisCO (Cre02.g120100 AND Cre02.g120150), while the chloroplast-encoded large subunit was omitted all together. This and other erroneous GPR rules were refined to make them consistent with experimental and sequence data (Table S4). In addition, 10 orphan gap-

filling reactions, which had no GPR associations in *iRC1080*, were assigned high-scoring candidate genes based on sequence analysis, while 10 reactions were orphaned (i.e., lost their GPR associations) as previously assigned genes did not have any literature- or sequence-based evidence for the assignment or could not be mapped to any loci in the new genome annotation (Table S4).

Previous analyses led to the generation of high-quality, experimentally derived biomass reactions for simulation of photoautotrophic (CO_2 as the sole carbon source in the presence of light), mixotrophic (CO_2 and acetate used as carbon sources in the presence of light) and heterotrophic (acetate used as carbon source in the dark) growth in *C. reinhardtii* (Boyle and Morgan, 2009; Chang et al., 2011). However, the maintenance requirements were not quantified and these values were assumed to be similar to those of other well-studied organisms (Boyle and Morgan, 2009). Subsequent experimental analysis of photosynthetic maintenance energy requirements indicated that these assumed values are likely too low (Kliphuis et al., 2011). Consequently, we updated the photosynthetic maintenance energy values of the photosynthetic biomass reactions from *iRC1080* with the experimentally determined growth- and non-growth-associated maintenance energy values (Kliphuis et al., 2011). Furthermore, several essential cofactors including NAD(H), NADP(H), FAD, biotin and thiamine, which were omitted from previous biomass reactions, were added to generate more comprehensive biomass reactions for simulating *C. reinhardtii* growth (Table S5).

iCre1355: a refined and up-to-date genome-scale reconstruction for *C.*

reinhardtii—Combining the above-described 312 metabolic genes and their associated reactions with the refined metabolic components from *iRC1080*, we generated a new comprehensive genome-scale reconstruction for *C. reinhardtii*, *iCre1355* (Table S5–S7). This reconstruction consists of 1,355 genes and 1,460 transcripts, accounting for ~7.5% of the annotated genes and transcripts in the *C. reinhardtii* genome (Table 1), an increase of ~26% over *iRC1080*. *iCre1355* also contains a total of 2,394 reactions (including 1,862 transformation reactions) and 1,133 unique metabolites (Table S5 and S6), encompassing a broad range of metabolic functions (Figure 1c). Similarly to *iRC1080*, reactions in *iCre1355* are distributed over 9 intracellular compartments and the extracellular space, with the majority of reactions localized to the cytosol, chloroplast and mitochondria (Figure 1d). Of the 2,394 reactions in *iCre1355*, 595 (24.9%) were blocked under the conditions tested and these reactions are distributed over a wide variety of subsystems (Figure 1e). Of these blocked reactions only 82 (13.8%) were associated with the 312 newly incorporated genes. From a global analysis of essential genes, we found that 250 (18.5%), 161 (11.9%) and 173 (12.8%) of the genes in *iCre1355* were predicted to be essential for photoautotrophic, mixotrophic and heterotrophic growth, respectively (Figure 1f), with 131 (9.7%) of these genes essential for growth under all 3 conditions, representing the core set of predicted essential genes in *iCre1355*.

Pathways in *iCre1355* have characteristic expression profiles—While all the genes included in *iCre1355* have been manually curated and assigned their known or most likely biological function, it is also of value to ensure that these genes are actively expressed within the cell, at least under a subset of relevant conditions. A relatively large number of

transcriptomics studies have been conducted recently that provide systems-level insights into aspects of *C. reinhardtii* physiology. These gene expression datasets were used to study processes ranging from the cellular response to nutrient depletion to oxidative stress. In order to assess the expression of genes in the various pathways included in iCre1355, we downloaded publicly available processed RNA-seq datasets from Phytozome 10.1, which were obtained from 6 studies (Castruita et al., 2011; Boyle et al., 2012; Urzica et al., 2012; Urzica et al., 2012; Duanmu et al., 2013; Hemschemeier et al., 2013). We then used the fraction of highly expressed genes (the top 5% most highly expressed genes per experiment) within each pathway as a metric for assessing the level of expression of pathways across conditions (Figure 2). From this analysis, we observed a large amount of condition-dependent change in pathway-level expression, much of which was consistent with known expression patterns that arise in response to changes in the environment (Figure 2). For example, changes such as the decrease in mRNA levels of the chlorophyll biosynthetic pathway during nitrogen starvation (Boyle et al., 2012), the increased mRNA abundance of ascorbate metabolism genes during oxidative stress with H₂O₂ (Urzica et al., 2012) and the increase in oxidative stress response (glutathione and ascorbate metabolism pathways) during transition from dark to light (and its attenuation by biliverdin) (Duanmu et al., 2013), amongst others, can be easily discerned. Nevertheless, the overall profile of expressed genes in each pathway was sufficiently consistent across all conditions in this dataset such that the pathways could be partitioned into 4 distinct groups: low, moderate, high and very high expression clusters (Figure 2). As might be expected, the very highly expressed pathways included central processes such as photosynthesis, TCA cycle and oxidative phosphorylation, while the highly expressed pathways included many growth-associated pathways such as amino acid and fatty acid metabolism amongst others. On the other hand, the lowly expressed pathways included those involved in cofactor biosynthesis, sphingolipid metabolism and detoxification processes, functions that are likely not required at high levels under most growth conditions. Using low expression thresholds of between 0.5 and 5 fragments per kilobase of transcript per million mapped reads (FPKM) (Kellis et al., 2014), the pathways in iCre1355 had between 98.3% and 77.6%, of their genes expressed above these thresholds, respectively, under at least one condition. At an intermediate threshold of 3 FPKM, greater than 80% of the genes in all pathways were expressed under at least one condition except for the callose biosynthesis, sulfur metabolism, nitrogen metabolism and polyamine metabolism pathways, where only between 50% and 77% of the genes were expressed above 3FPKM across all conditions (Figure S2).

Evaluation of iCre1355 demonstrates significant improvements in prediction accuracy

To assess the predictive accuracy of the constraint-based model (CBM) derived from iCre1355, we evaluated its predictions against experimental data obtained from genotype-phenotype analysis and growth rates of cells cultured in chemostats.

iCre1355 accurately predicts gene deletion phenotypes—A standard approach for assessing the predictive performance of constraint-based models is through comparison of model predictions to experimentally determined gene deletion phenotypes (Oberhardt et al., 2009). Unlike model organisms like *Escherichia coli* and *Saccharomyces cerevisiae*, *C. reinhardtii* does not have a comprehensive catalogue of known gene deletion phenotypes.

Thus, to generate a list of genotype-phenotype relationships for model evaluation, we conducted a thorough review of primary literature and the Chlamydomonas Resource Center (<http://chlamycollection.org/strains/>) collection of mutants. This led to the generation of a list of 81 well-defined *C. reinhardtii* mutants of metabolic genes with known phenotypes under photoautotrophic, mixotrophic and heterotrophic conditions (Table S8). When combined with media supplementation rescue phenotypes, this led to the generation of a total of 306 genotype-phenotype relationships for model evaluation (Table S9).

We used the CBM derived from *iCre1355* to predict the phenotypes of these 81 mutants under photoautotrophic, mixotrophic and heterotrophic conditions and compared these to experimental data. We also evaluated the predictions of *iRC1080* under the same conditions, but for 48 mutants (corresponding to 183 genotype-phenotype relationships), as only this subset were represented in that model (Table S8). This analysis showed that *iCre1355* allows for very accurate predictions of gene deletion phenotypes, with an area under the receiver operating characteristic curve (AUC) of 0.92 (with a precision of 80%, sensitivity/recall of 83%, specificity of 92% and an accuracy of 90%) (Figure 3a). On the other hand, the 48 mutants evaluated with *iRC1080* resulted in an AUC of 0.825 (with a precision of 54%, sensitivity/recall of 62%, specificity of 84% and an accuracy of 79%), indicating that *iCre1355* is a better predictor of genotype-phenotype relationships, at least for the dataset used in this analysis. We also assessed *iCre1355*'s predictions on the same set of 48 mutants as that used to assess *iRC1080*. For this dataset, *iCre1355* showed an AUC of 0.9 (with a precision of 67%, sensitivity/recall of 90%, specificity of 88% and an accuracy of 88%) (Figure 3a). To investigate the impact of the modified biomass reactions on the predictions of *iCre1355*, we made an independent set of predictions using the original biomass reactions from *iRC1080* instead of those developed in this study. Assessing predictions for the consequences of mutations in 48 genes present in both models, we found that *iCre1355* had an AUC of 0.84 (with a precision of 63%, sensitivity/recall of 67%, specificity of 89% and an accuracy of 84%) (Figure 3a). Given the larger number of genes and reactions that have to be evaluated when assessing *iCre1355* and the fact that the AUC does not normalize for the total number of genes evaluated (directly or indirectly via associated GPR rules), the 1.5% increase in AUC for this evaluation could in fact be larger. Thus the improved predictive power of *iCre1355* is due to a combination of model additions, model refinements and refined biomass reactions.

***iCre1355* accurately predicts maximum triacylglycerol yields**—An interesting characteristic of *C. reinhardtii* and other microalgae is their ability to accumulate large stores of neutral lipids such as triacylglycerols (TAGs), particularly when faced with environmental stresses that inhibit growth (Chisti, 2007; Scott et al., 2010; Boyle et al., 2012). This trait has been intensely studied due to its potential biotechnological value. Thus, it is important that a model of *C. reinhardtii* metabolism accurately predicts TAG yields. We assessed the predicted maximum yield of TAG on CO₂ and light (photons) using *iCre1355* and *iRC1080* under photoautotrophic conditions, and found that it was identical for both models (at ~0.36 g TAG/g CO₂). This value is also identical to the calculated theoretical maximum value (Figure 3b) (see Experimental procedures). This indicates that the

stoichiometry of the reactions leading from CO₂ to TAG biosynthesis in both models is accurate.

Conversely, while the predicted maximum TAG yield on light by *iCre1355* was 1.43 g TAG/mol photon, which is close to the calculated theoretical maximum of 1.38 g TAG/mol photon (see Material and methods), that predicted using *iRC1080* was significantly higher at 1.78 g TAG/mol photon (Figure 3b). The slight difference between the *iCre1355*-predicted and the calculated maximum TAG yield likely reflects the uncertainty in the rotational fold symmetry (or H⁺/ATP ratio) of the *C. reinhardtii* F1F0 ATP synthase (i.e., the number of protons required for one complete rotation of the ATP synthase to produce 3 ATP molecules, which is dependent on the number of identical c subunits (Stock et al., 2000)). Using values between 14 (as observed in spinach chloroplasts (Seelert et al., 2000)) and 12 (as observed for *E. coli* (Jones et al., 1998)) resulted in predicted maximal yield values between 1.33 and 1.43 g TAG/mol photon. A value of 13, which has been observed in cyanobacteria (Pogoryelov et al., 2007), resulted in a yield closest to the calculated theoretical maximum. Changes in this value did not alter predictions from *iRC1080*, indicating that the predicted yield values were independent of ATP generated via ATP synthase. This erroneous prediction of *iRC1080* was due to the presence of cycles involving ATP-generating reactions in *iRC1080* that result in production of unlimited amounts of ATP independent of the ATP synthase. Specifically, the pyruvate carboxylase reaction, which is irreversible under physiological conditions (Nelson and Cox, 2005), and ATP:GTP 3'-diphosphotransferase reaction are both reversible in *iRC1080* and this results in the generation of ATP independent of ATP synthase. These reactions were made irreversible in *iCre1355*. In addition to resulting in erroneous TAG yield predictions in *iRC1080*, these loops also negatively impact gene essentiality predictions, as the chloroplast ATP synthase is not predicted to be required for growth under photoautotrophic conditions, contrary to experimental observations. While constraint-based models can be powerful tools to study cellular metabolism on a large-scale, they typically contain a large number of futile cycles, many of which may not be biologically relevant to due spatial, temporal or other regulatory constraints. Thus, analyzing models for such cycles, and eliminating them where necessary, can be of value for specific applications.

***iCre1355* allows for accurate growth rate predictions**—To assess the ability of *iCre1355* to predict growth rates, we obtained data on macronutrient uptake rates from *C. reinhardtii* cells grown in continuous culture under photoautotrophic, mixotrophic and heterotrophic conditions. Uptake rates of carbon (either as dissolved inorganic carbon [DIC] or acetate), nitrogen (NH₄) and phosphorus (PO₄) were measured (Table 2) and used as initial input parameters for predicting growth rate with *iCre1355*. Comparing the measured chemostat growth rates to the *iCre1355* FBA-predicted growth, when using all 3 measured uptakes rates as modeling constraints, showed good overall agreement between the predicted and observed values, with an R² of 0.83 (p-value = 0.0002) (Figure 3c). These observations suggest that *iCre1355* accurately predicts *C. reinhardtii* growth rates even with minimal input data.

To determine how efficiently chemostat-grown *C. reinhardtii* cells use each of the measured macronutrients, the uptake rate for each macronutrient was individually used as a constraint

in *iCre1355*. This analysis revealed that, in general, the carbon and phosphate uptake rates served as the most useful predictors of cellular growth rate across all 3 growth modes (Figure 3d). On the other hand, we consistently observed significantly higher nitrogen uptake rates than what would be predicted to be necessary of optimal biomass production (Figure S3). This might indicate some excess uptake of NH_4 during steady state growth in NH_4 -replete media, which could be redirected to internal nitrogen reserves. Alternatively, this apparent excess of consumed NH_4 could point to a nitrogen deficiency in the *iCre1355* biomass reaction definition. Further work will be required to identify the cause(s) of this discrepancy. Overall this analysis forms a basis for the use of *iCre1355* to study growth rate in *C. reinhardtii*.

Using *iCre1355* to study key physiological processes in *C. reinhardtii*

Having validated *iCre1355* at various levels, we then used it to study two key processes in *C. reinhardtii*: (i) the metabolic response to nitrogen starvation; and (ii) the effects of changing light regimes on growth. To achieve this, we leveraged available transcriptomics datasets. While transcriptome-level data only provides one layer of information about the complexities of regulation and information transfer, as metabolic adjustments can occur post-transcriptionally and post-translationally (Branco-Price et al., 2008; Krasensky and Jonak, 2012), integration of high-throughput genome-wide measurements with constraint-based models have previously been shown to be valuable for reducing the solution space of flux distributions, potentially improving predictive accuracy and allowing for new biological insight (Reed, 2012; Kleessen et al., 2015). In order to study the metabolic changes that result from shifts in environmental conditions, we used previously published high-resolution time-series transcriptomics datasets obtained from *C. reinhardtii* cells subjected to nitrogen starvation (Boyle et al., 2012) and changing light intensity (Mettler et al., 2014) as additional constraints on *iCre1355*.

Integration of these datasets into *iCre1355* was accomplished by combining two well-established approaches: E-flux (Colijn et al., 2009), which imposes gene expression data-derived bounds on the maximum flux through each reaction in the network, and iMAT (Shlomi et al., 2008; Zur et al., 2010), which maximizes the agreement between metabolic fluxes and reactions categorized as having a high or low activity based on transcriptomics data (see Experimental procedures). It should be noted that predictions derived for integration of transcriptomics data into metabolic models will not capture the lag in information transfer from transcription to synthesis of active enzymes and this needs to be considered during interpretation of results from integration of time series data.

The metabolic response of *C. reinhardtii* to nitrogen starvation

When faced with a variety of environmental stresses such as nutrient limitation, *C. reinhardtii* accumulates TAGs (Boyle et al., 2012; Urzica et al., 2013). This process has been extensively studied under a variety of micro- and macro-nutrient starvation conditions (Miller et al., 2010; Boyle et al., 2012; Blaby et al., 2013; Urzica et al., 2013; Iwai et al., 2014). In particular, the cellular response to nitrogen starvation has been analyzed in several studies using system-wide measurements in *C. reinhardtii* (Miller et al., 2010; Boyle et al., 2012; Blaby et al., 2013; Park et al., 2015) and other microalgae (Recht et al., 2014). In

particular, Boyle et al. (Boyle et al., 2012) analyzed the time course of the transition from nitrogen-replete to nitrogen-deplete media at very high resolution using RNA-seq analysis. While this dataset was not obtained from cells at growing at steady state (a key assumption of constraint-based analysis) due to the nature of the response to this environmental perturbation, it is likely that valuable information could still be gleaned from these data by incorporating them into a large-scale metabolic model. Thus, we used this dataset as constraints on *iCre1355* and analyzed the predicted effects on growth, TAG accumulation and pathway activity to gain a better understanding of the metabolic state of the cell during this transition.

Cessation of growth during nitrogen starvation—Applying these transcriptional constraints across the time course spanning 48 hours and maximizing for growth rate at each time point (E-flux), we observed a gradual decrease in predicted optimal growth rate (Figure 4a), with the slowest predicted growth rate at 48 hours. Given that maximization of growth is unlikely to be a relevant objective function during nitrogen starvation conditions, we extended the E-flux approach by employing an alternative objective function, i.e., maximization of the agreement between expression levels and metabolic fluxes (iMAT). With this approach, we used E-flux to set “hard constraints” on the upper and lower bounds on each reaction, while iMAT was used to impose “soft constraints” on which reactions carry flux. Using this combined approach, the predicted growth rate showed a similar trend, however, by 8 hours, a complete cessation of growth is predicted (Figure 4a) that is maintained until the end of the time course. These predictions are consistent with the experimentally observed cessation of growth during nitrogen starvation, wherein *C. reinhardtii* cells are typically observed to undergo 1 to 2 doublings before a halt in growth (Boyle et al., 2012). The predicted halt in growth at 8 hrs by E-flux+iMAT also coincides with the first time point at which significant TAG accumulation is observed (Boyle et al., 2012). It should be noted that only transcriptomics-based constraints were applied for these simulations, while nutrient and light uptake rates were left unconstrained. Nevertheless, the predicted growth rate at the 0 minute time point ($\sim 0.06 \text{ h}^{-1}$) is well within the range of mixotrophically grown *C. reinhardtii* cells (Table 2) (Mettler et al., 2014), while the predicted uptake rates of acetate ($3.9 \text{ mmol gDW}^{-1} \text{ h}^{-1}$), NH_4 ($0.31 \text{ mmol gDW}^{-1} \text{ h}^{-1}$) and PO_4 ($0.047 \text{ mmol gDW}^{-1} \text{ h}^{-1}$) are also within range of the experimentally measured values (Table 2). These observations suggest that transcriptional constraints can be sufficient for making accurate growth rate predictions, in lieu of measured uptake rates.

Transcriptome-predicted TAG production potential is highest in early stages of nitrogen starvation—A similar analysis for TAG production predicted that early stages of the nitrogen starvation response have the greatest potential for TAG production. Maximizing for TAG production using E-flux reveals a gradual increase in TAG production potential starting at 4 minutes after nitrogen depletion and reaching a maximum value at 30 minutes before a gradual decline (Figure 4b). This suggests that while significant TAG accumulation is not observed until 4 hours after the onset of nitrogen starvation, *C. reinhardtii* cells may begin preparations for this state transition much sooner. Eflux+iMAT also predicted a similar trend, although the predicted peak is very early (12 minutes) and the

predicted TAG flux is about an order of magnitude smaller than when maximizing for TAG production.

Nitrogen starvation results in altered pathway activity levels—Stress induced by nitrogen starvation results in systems-level changes in gene expression and metabolism (Miller et al., 2010; Boyle et al., 2012; Park et al., 2015). To gain insight into these global changes in metabolism, we assessed the transcriptome-predicted changes in reaction activities across the various pathways included in *iCre1355*. By applying thresholds that enabled division of genes into high and low expression groups, we combined transcriptomics data with *iCre1355*'s GPR rules to categorize the model reactions into “high” and “low” activity states (see Experimental procedures). We then assessed how nitrogen starvation causes pathway-level changes in activity by using changes in the number of highly active reactions as a measure of overall change pathway activity.

From this analysis, we observed that pathways could largely be categorized into 3 groups (Figure 5). Group I (growth-associated pathways) consists of pathways for which the constituent reaction activity levels were predicted to gradually decrease over the time course of the nitrogen starvation experiment (Figure 5). Unsurprisingly, this group included several pathways involved in biosynthetic processes such as amino acid biosynthesis, purine and pyrimidine biosynthesis and cofactor biosynthesis. In addition, consistent with results from previous reports, reactions involved in porphyrin and chlorophyll biosynthesis, as well as photosynthesis, also showed significant reduction in predicted activity during nitrogen starvation. Despite the observed accumulation of TAGs during nitrogen starvation, fatty acid and glycerolipid biosynthetic pathways showed a significant reduction in transcriptome-predicted activity, suggesting that *de novo* fatty acid biosynthesis is reduced during starvation. This is consistent with previous reports indicating that recycling of membrane fatty acids may make a significant contribution of fatty acids used for TAG accumulation (Fan et al., 2011), limiting the need for *de novo* fatty acid biosynthesis. The lowering of activity of all these biosynthetic pathways is also consistent with the cessation of growth during nitrogen starvation.

Group II (stress response and carbon storage pathways) consists of pathways that showed an initial increase in predicted activity in the early stages of nitrogen starvation (within the first 30 minutes), followed by a gradual decrease in predicted activity (Figure 5). This group included several cofactor/vitamin biosynthesis pathways (e.g., folate biosynthesis, pantothenate/CoA biosynthesis and ascorbate biosynthesis), and the carotenoid biosynthetic pathway. Increase in ascorbate, folate and other antioxidant biosynthetic genes has previously been linked to the response to oxidative stress in *C. reinhardtii* (Mittler, 2002; Muller-Moule et al., 2003; Urzica et al., 2012), while carotenoids are known to provide protection to the cell from reactive oxygen species that arise as a consequence of oxygenic photosynthesis (Muller-Moule et al., 2003; Perez-Perez et al., 2012). The predicted transient increase in the activity of these pathways could indicate a state of increased oxidative stress in the early stages of N-starvation, possibly resulting from an excess in reducing power produced during photosynthesis, which can no longer be channeled to biomass as growth rate reduces. In contrast to most other amino acid biosynthetic pathways, which are found in Group I, histidine biosynthetic and glycine/serine/threonine biosynthetic pathway activities

are predicted to increase transiently early in N-starvation. An increase in glycine biosynthetic genes in response to starvation has also been observed in yeast cells and could also be linked to the oxidative stress response (Petti et al., 2011).

Starch metabolism is predicted to increase in activity early in nitrogen starvation before a gradual decrease, consistent with previous reports (Boyle et al., 2012; Blaby et al., 2013; Park et al., 2015). A similar pattern is also observed with glycerophospholipid metabolism pathway. The early activation of these pathways may prime cells for storage of starch and lipids later during the course of N-starvation. The early increase in the predicted activity of the glycerophospholipid pathway is also consistent with the predicted early increase in TAG production capacity (Figure 4b).

Group III (stress response and nitrogen scavenging pathways) consists of pathways that showed a predicted increase in activity into the later stages of N-starvation. This included the transport reactions, the pentose phosphate pathway, nitrogen metabolism and the arginine biosynthetic pathway. The increase in NH_4 and related transporters in response to nitrogen starvation has previously been reported and our analysis also revealed a similar trend, with a wave of transporters being induced sequentially from the onset of N-starvation (Figure 5). First, the polyamine (putrescine and spermidine) transporters become highly active, followed by the NH_4 transporters, uric acid transporters, and then a series of transporters involved in the transport of amino acids between cellular compartments and/or from the environment. These likely represent mechanisms used by the cells in an attempt to scavenge any nitrogen-containing compounds in their environment, as well as shuttle amino acids to compartments where they are most required. The predicted increase in the nitrogen metabolism pathway is attributable to the nitric oxide oxidoreductase (Cre01.g000350), which catalyzes the conversion of NO to NO_3 , and glutamate synthase (Cre13.g592200), which catalyzes the conversion of glutamine and α -ketoglutarate to two molecules of glutamate. The cells could potentially use these reactions as routes for consuming excess reducing power from photosynthesis, as well as generating usable nitrogen sources. Increased glutamate synthase enzyme activity has been observed in nitrogen-starved *C. reinhardtii* cells (Menacho and Vega, 1989). Nitric oxide oxidoreductase may also act to attenuate NO signaling, which inhibits the high-affinity NH_4 uptake induced during nitrogen starvation (Sanz-Luque et al., 2013).

Our analysis also indicates that a portion of the arginine biosynthetic pathway increases in activity in response to nitrogen starvation. This included 7 reactions required for the conversion of glutamate into ornithine, potentially linking the product of the glutamate synthase reaction to arginine biosynthesis. An increase in arginine biosynthetic genes was also reported in (Park et al., 2015), however no accumulation of intracellular arginine or ornithine was observed.

Assessing the number of lowly expressed reactions per pathway across time also provides an interesting complementary view to the changes in pathway activity linked to N-starvation (Figure S4). Overall, these analyses provide a comprehensive overview of the pathway-level changes that occur during the metabolic response to nitrogen starvation.

The metabolic response of *C. reinhardtii* to changing light intensity

To thrive in nature, photosynthetic microbes need to respond appropriately to changes in light intensity, which can occur over a wide range of time scales from a few seconds (such as shading events) to days (such as diurnal cycles) or months (seasonal cycles) (Mettler et al., 2014). Changes in light intensity directly impact the growth of phototrophs, affecting their ability to fix CO₂ and thus requiring photosynthetic cells to adjust their metabolism. To assess the changes that occur in the transition from light-limiting conditions to higher light intensities, Mettler et al. (Mettler et al., 2014) obtained global mRNA abundance measurements from *C. reinhardtii* cells growing at steady state at 0, 40, 120 and 480 minutes after a switch from low light (41 μmol photons m⁻² s⁻¹) to a higher light intensity (145 μmol photons m⁻² s⁻¹). We used these data as constraints on *i*Cre1355 to study how transcriptional changes affect flux distributions during this transition.

Transcriptional constraints alone are insufficient to predict differences in growth due to changing light intensity—Applying these transcriptional constraints across the time course and maximizing for growth rate (E-flux) or agreement between transcriptome data and flux (E-flux+iMAT), we observed very little difference in the predicted growth rate between conditions (Figure S5a). While the predicted growth rates between 40 and 480 minutes were very similar to the reported experimental growth (0.08 h⁻¹) for both approaches, the predicted growth rate at low light (0 m) was much higher than the reported value of 0.02 h⁻¹. This suggests that the subtle transcriptional changes that occur during this transition were insufficient to constrain the model to permit accurate prediction of growth rate. Thus, to assess changes in metabolic flux distribution associated with light-limited growth, we additionally adjusted the photon uptake rate to a value that resulted in a predicted growth rate of 0.02 h⁻¹ (~8 mmol photon/gDW⁻¹ h⁻¹).

Changes in light intensity result in distinct temporal changes in flux capacity across pathways—Using a similar approach to that used to study pathway level changes in activity during nitrogen starvation; we assessed the predicted changes in pathway activity resulting from a change in light intensity. Given the relatively subtle changes in gene expression during this transition, only a few pathways were predicted to show significant changes in reaction activity via this approach (Figure S5b). The greatest observed changes were for pathways involved in fatty acid and nucleotide biosynthesis, which increased upon changing from low to higher light, consistent with cells increasing in growth rate, as these macromolecules are major components of biomass.

To gain a more detailed view of the potential impact of transcriptional changes on the metabolic state of the cell during this transition, we assessed the differences in predicted flux capacity across conditions using flux variability analysis (Mahadevan and Schilling, 2003). With transcriptomic constraints imposed on the CBM, we determined the minimum and maximum allowable flux through each reaction in the network. The flux capacity through each reaction was then used to explore the effects of changing light intensity on potential pathway usage (see Experimental Procedures). A comparison of the correlation of flux capacities across conditions shows that the higher light conditions are more similar to each other than to the low light condition (Figure S5c). This is consistent with the equivalent

growth rate of the cells at higher light (0.08 h^{-1}), which is four times faster than at time 0 minutes. It should be noted that a similar correlation analysis using raw gene expression data for the metabolic genes in *iCre1355* did not group higher light experiments together particularly well (Figure S5c), consistent with a previously conducted principal component analysis (Mettler et al., 2014). This suggests that the predicted flux capacity might be a useful metric to assess transcriptome-predicted changes in metabolic activity.

Assessing the changes in flux capacity on a pathway level, we observed that the pathways could in general be divided into four broad categories based on how and when they responded to changes in light intensity. Group I was made up of pathways wherein reactions showed an increased flux capacity by the first higher light time point (40 minutes) compared to the initial time point (0 minutes) and remained high throughout the rest of the time course (Figure 6). Unsurprisingly Group I included many pathways associated with growth including amino acid biosynthesis, fatty acid and lipid biosynthesis, carbon fixation, central metabolism, nucleotide biosynthesis and photosynthesis-related pathways. These appear to be the “first responders” to increased light intensity pointing to their important contribution to biomass composition, maintenance of redox balance and bioenergetics during photosynthetic growth. The predicted quick response of these pathways is also consistent with metabolite data collected by Mettler et al. (Mettler et al., 2014), which shows metabolites involved in the Calvin-Benson cycle carbon fixation reactions and central carbon metabolism increasing rapidly on the switch to higher light intensity. Of the 17 metabolites observed to show statistically significant increases in abundance (Mettler et al., 2014), and which are present in *iCre1355*, 15 were metabolic intermediates in pathways grouped among the “first responders” including carbon fixation intermediates (5 metabolites), glycolysis intermediates (3 metabolites), amino acids (5 metabolites), TCA cycle intermediates (1 metabolite) and nucleotide sugar (1 metabolite). These predictions are also consistent with metabolite data collected from *C. reinhardtii* cells subjected to high light stress, as well as those acclimated to high light conditions (Davis et al., 2013), which show significant increase in amino acid pools that align well with the predicted pathway activities.

Group II pathways did not show an increase in flux capacity until the 120 minutes or 480 minutes time point. These “late responders” consist of pathways for the biosynthesis of amino acids such as lysine and histidine, which make up a relatively small proportion of amino acids found in *C. reinhardtii* proteins (Boyle and Morgan, 2009). This group also included several pathways involved in vitamin/cofactor biosynthesis, which are also typically required in small quantities, thus their delayed response might be expected, as these metabolites may not become limiting until a substantive amount of dilution by cell division has occurred.

Group III pathways showed an initial decrease in flux capacity at 40 minutes, but an increase in flux capacity by 120 minutes (or 480 minutes in the case of pyruvate and propanoate metabolism). These can also be thought of as “late responders”. In addition to pathways in these three categories, a fourth group consisting of 20 pathways showed no change in flux capacity relative to 0 minutes. These pathways were broadly distributed across subsystems. Overall, these analyses provide insights into the time-dependent response

and usage of pathways, as photosynthetic cells adjust their metabolism in response to increased light.

Conclusions

Microalgae will likely play a pivotal role as we strive to build more sustainable societies due to their ability to produce a suite of chemicals of socio-economic value with low environmental impact. *C. reinhardtii*, being the best characterized microalga, will undoubtedly play a leading role in informing genetic strategies aimed at the rational design of algal strains with improved properties. In this study, we conducted a systems-level analysis of *C. reinhardtii* metabolism beginning with the assembly of a refined genome-scale metabolic reconstruction *iCre1355*. Our assessment and experimental validation of the predictions from *iCre1355* highlight its predictive accuracy and its potential value as a tool for generating hypothesis to guide future experimental design. We further illustrated the value of *iCre1355* through the integration of disparate gene expression datasets, combining analyses of network structure with constraint-based approaches to gain systems-level insights into the metabolic responses of *C. reinhardtii* to changing environmental conditions. Our analysis of the response to nitrogen starvation yielded several predictions that were consistent with previous analyses, providing an additional level of validation for *iCre1355*. Furthermore, this analysis pointed to an early concerted response to oxidative stress potentially faced by the starved cells as photosynthesis continues despite a reduction in growth rate. In addition, the early priming of the starved cells for both starch and lipid storage is predicted, though starch storage typically precedes lipid storage. Our analysis of the metabolic response to increasing light intensity identified the temporal response of pathways, with the “first responder” pathways being involved in the synthesis of metabolites that make large contributions to overall biomass, consistent with the observed increased growth rate. Overall, the improved genome-scale metabolic reconstruction and comprehensive analysis conducted in this study provides an enhanced framework for future systems-level analysis of *C. reinhardtii* and microalgae metabolism.

Experimental procedures

Strains and growth conditions

C. reinhardtii strain CC-1690 obtained from the Chlamydomonas Resource Center (<http://chlamycollection.org>) was used for all growth experiments. Starter cultures for the bioreactors were grown in shaker flasks at 22° C at 120 rpm, illuminated with cool white light bulbs at ~80 $\mu\text{mol photons m}^{-2} \text{ s}^{-1}$. Photoautotrophic cultures were grown at ambient CO₂ in modified Artificial Seawater Media (MASM) (NH₄Cl – 7.5mM, MgSO₄ – 10mM, KCl – 8.05mM, CaCl₂.H₂O – 2.04mM, KH₂PO₄ – 0.37mM, tris base – 8.25mM, trace elements (Na₂EDTA.2H₂O – 0.0161mM, FeCl₃.6H₂O – 0.0043mM, MnCl₂.4H₂O – 0.0022mM, ZnSO₄.7H₂O – 0.00092mM, Na₂MoO₄.2H₂O – 0.00031mM, CoCl₂.6H₂O – 0.0001mM)), while mixotrophic and heterotrophic cultures were grown in tris acetate phosphate medium (TAP). For heterotrophic cultures, the acetate concentration was reduced from ~17 mM to ~10.625 mM to prevent any growth inhibition (Chen and Johns, 1994).

Bioreactor setup

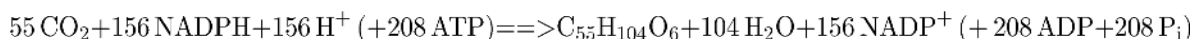
C. reinhardtii cells were grown in continuous culture using a New Brunswick BioFlow/CelliGen 115 bioreactor (Eppendorf inc). For photoautotrophic experiments, cells were grown on MASM and bubbled with air enriched with 0.2% CO₂. For mixotrophic experiments, cells were grown on TAP and bubbled with air enriched with 0.2% CO₂, while for heterotrophic experiments, cells were grown on TAP and bubbled with air. All chemostats were agitated at 75 rpm. Photoautotrophic and mixotrophic experiments were illuminated at ~150 μmol photons m⁻² s⁻¹ with cool white light bulbs. For heterotrophic experiments, the bioreactor was wrapped in foil to keep it in the dark. Independent one-liter reactors were set up for each biological replicate, which were initially started in batch mode and run for 2–5 days (depending on growth condition) until the cell density within the bioreactor was ~1×10⁶ cells/mL. The bioreactors were then switched to continuous mode at the target dilution rate (Table 2). Bioreactors were run in continuous mode for 4 to 5 chemostat volumes (4 to 5 L) when steady state was typically achieved. Steady state was assumed when constant cell counts were achieved over a 24- to 48-hour time period. Samples were sterilely taken from the bioreactor with a syringe from the sampling port to keep track of cell counts.

Uptake rate measurement

When steady state was achieved, the chemostats were harvested to quantify cell dry weight and nutrient uptake. The concentrations of macronutrients left over in the filtered media from each chemostat were determined and compared to the concentrations in the starting media to determine uptake rates. NH₄ concentrations were determined using the ammonia assay kit (AA0100; Sigma-Aldrich), phosphate concentrations were determined using the SensoLyte MG colorimetric phosphate assay kit (AS-71103; AnaSpec) and acetate concentrations were determined using the acetate colorimetric assay kit (MAK086; Sigma-Aldrich). Dissolved inorganic carbon (DIC) was assayed using a DIC Analyzer (Apollo Scitech Inc) attached to a LI-7000 Gas Analyzer (LI-COR). To determine, DIC uptake rates, chemostats were run in batch mode without bubbling for ~90 minutes after initial harvesting of samples for uptake rate measurement. Samples were then taken at ~15 minute intervals and analyzed for a change in DIC concentration. Changes in the DIC concentration over these time intervals were averaged to determine the DIC uptake rates per hour. Five to fifteen milliliters of cells (depending on chemostat cell density) were harvested to determine the chemostat cell dry weight. The ash free dry weight was calculated by subtracting 5.68% from the measured dry weight based on previous analysis of ash content of *C. reinhardtii* cells (Kliphuis et al., 2011).

Yield calculations

TAG yields were calculated using one of the most abundant TAG species observed in *C. reinhardtii* cells (tag16018111Z180 – 16:0/18:1(11Z)/18:0, see Table S5). The balanced net reaction for the formation of this TAG species would be:



This reaction accounts for all the reducing power consumed (in the form of NADPH) and produced (in the form of NADH, which is assumed to be converted to NADPH) during the formation of one molecule of TAG. To calculate the theoretical maximum TAG yield on CO₂ uptake, we can see from the reaction equation that 1 mol of CO₂ would generate 0.018182 mol of tag16018111Z180 (i.e., 1/55). Thus, given the molecular weights of CO₂ (44 g) and tag16018111Z180 (860 g), the theoretical maximum TAG yield on CO₂ would be 0.355375 g TAG/g CO₂. To calculate the theoretical maximum TAG yield per mol photon, we followed the established stoichiometry of photochemistry i.e., 8 photons results in the production of 2 molecules of NADPH (and 3 molecules of ATP). Based on this, 624 photons would be required to produce the 156 molecules of NADPH needed for the synthesis of 1 mol of tag16018111Z180. Thus, given the molecular weight of tag16018111Z180 (860 g), the theoretical maximum TAG yield on photons would be 1.378 g TAG/mol photon.

Metabolic network reconstruction

Candidate identification and manual curation—The previously reconstructed genome-scale model for *C. reinhardtii* iRC1080 served as the reference point for model reconstruction. The process of model refinement began with the extraction of all *C. reinhardtii* specific metabolic and genomic information from four public databases: the Plant metabolic network (ChlamyCyc 4.0), KEGG (Kanehisa et al., 2002; Kanehisa et al., 2014), NCBI and Phytozome 10.1. ChlamyCyc 4.0 and KEGG served as the main data sources for metabolic information, NCBI for the chloroplast and mitochondrial genomic information, while the most up-to-date nuclear genomic information was obtained from Phytozome 10.1 (v5.5 as of this writing). This process led to the identification of 586 candidate metabolic genes, which could not be mapped to any genes in iRC1080 and served as the initial set of genes for manual curation. The process of manual curation involved: (i) primary literature searches for evidence a candidate gene performs the assigned function; (ii) functional domain analysis on the predicted protein sequence of candidate genes using Pfam (Bateman et al., 2004) and BLAST to verify possession of appropriate functional domains to catalyze assigned reactions, as well as to verify GPR relationships obtained from ChlamyCyc 4.0; (iii) assessing connectivity of candidate genes (through its associated reaction(s)) to the rest of the network, with disconnected genes omitted from consideration). These curation steps resulted in the generation of a high confidence list of 312 candidate genes (320 transcripts) and 153 associated transformation reactions (Table S1).

Determination of subcellular localization—The subcellular localizations of candidate genes and reactions were determined based on the following criteria: (i) experimental evidence in literature from compartment specific proteomics analysis for the mitochondria (van Lis et al., 2003; Cardol et al., 2005; Atteia et al., 2009), chloroplast (Terashima et al., 2010; Terashima et al., 2011), or flagella (Pazour et al., 2005); or (ii) the consensus predictions of two subcellular localization prediction algorithms (TargetP (Emanuelsson et al., 2007) and Predotar (Small et al., 2004)). In the absence of evidence from either of these sources, genes and their associated reactions were assigned to the cytosol. The reactions were manually balanced for charge and mass, while reaction directionality was assigned based on heuristic rules (Henry et al., 2009) where applicable, or were made reversible.

Updating components from iRC1080—Genes from *iRC1080* in JGI v4 Aug5 annotation of the *C. reinhardtii* genome were mapped to JGI v5.5 using the algal functional annotation tool (<http://pathways.mcdb.ucla.edu/algal/index.html>) and these mappings were manually curated to correct erroneous mappings. Of the 1086 transcripts in *iRC1080* only 1062 could be mapped to the JGI v5.5 annotation after manual curation. The remaining 24 transcripts (12 genes), which could not be mapped to any JGI v5.5 annotation, were dropped from the model (Table S4). We applied similar manual curation steps as described for the 312 newly identified candidate genes to genes and reactions included in *iRC1080*. Based on these, 19 genes with incorrect functional assignments were dropped, 12 reactions lacking evidence were removed and several erroneous GPR rules for several reactions were corrected (Table S4). *PsbW*, an experimentally characterized nuclear-encoded gene (Bishop et al., 2003), is currently not annotated in JGI v5.5. *PsbW* was retained in *iCre1355* using its gene symbol (Table S5).

Gap filling—Gap filling was done manually and only when required for model functionality. Overall only 3 additional strictly gap filling reactions (for which no GPRs could be assigned) were added to *iCre1355*: MALCOAMT, PMEACPE, DBTS. Each of these reactions occurs in the biotin biosynthesis pathway. *C. reinhardtii* is capable of *de novo* biosynthesis of biotin (Croft et al., 2006) and the enzymes catalyzing the reactions upstream and downstream of these gap filling reaction are known or can be inferred from genomic information. Thus, these 3 reactions were included in *iCre1355* to complete the biotin biosynthetic pathway.

Biomass reaction and maintenance energy—The biomass reaction from *iRC1080* is very detailed and based on experimental data, thus we used these biomass reactions as the starting point for refinement. The following modifications were made to the *iRC1080* biomass reaction: (i) Growth and non-growth maintenance energy requirements (i.e., GAM and NGAM respectively) were based on assumed or fitted values (Boyle and Morgan, 2009). Subsequent efforts to determine maintenance energy requirements under photosynthetic conditions indicated that this assumption was far from accurate. Thus, maintenance energy values of the photoautotrophic and mixotrophic biomass reactions were modified to the experimentally determined values (GAM – 92.43 mmol ATP gDW⁻¹, NGAM – 2.85mmol ATP gDW⁻¹ h⁻¹), while the heterotrophic maintenance energy values from *iRC1080* were maintained; (ii) co-factors necessary for growth including NAD(H), NADP(H), FAD, glutathione, biotin and thiamine were included in each of the 3 biomass reactions in trace amounts (Table S5).

Constraint-based modeling

The genome-scale metabolic reconstructions *iRC1080* and *iCre1355* were used to generate separate stoichiometric matrices $S_{m \times n}$, where the rows, *m*, represent metabolites, and the columns, *n*, represent reactions in each reconstruction. The entries in the matrices represent the stoichiometric coefficients for metabolites involved in each reaction. *In silico* growth was predicted using FBA by solving the linear programming problem:

$$\begin{aligned} & \max \mathbf{v}_{\text{obj}} \\ & \text{s. t} \\ & S\mathbf{v}=0 \\ & \mathbf{V}_{\min} \leq \mathbf{v} \leq \mathbf{V}_{\max} \end{aligned}$$

where S is the stoichiometric matrix, \mathbf{v}_{obj} is the flux through the objective function (typically one of the 3 biomass reactions); \mathbf{v} is the vector of steady state reaction fluxes; and \mathbf{V}_{\min} and \mathbf{V}_{\max} are the lower and upper bounds for each flux, respectively. For reversible reactions, \mathbf{V}_{\min} and \mathbf{V}_{\max} were set to -100 and $100 \text{ mmol gDW}^{-1} \text{ h}^{-1}$, respectively, while for irreversible reactions \mathbf{V}_{\min} and \mathbf{V}_{\max} were set to 0 and $100 \text{ mmol gDW}^{-1} \text{ h}^{-1}$, respectively. In general, exchange reactions were made irreversible to only permit secretion of metabolites. However, exchange reactions for known media components were either set to measured values (e.g., for measured DIC, NH_4 , PO_4 or acetate uptake rates) or allowed to permit free exchange with the extracellular space (i.e., $-100 \leq \mathbf{v} \leq 100$). Simulations were conducted in the GAMS programming environment (GAMS Development Corporation, Cologne, Germany) using the CPLEX solver. *iCre1355* in GAMS format along with specific modeling constraints used to simulate growth under each condition are provided in Dataset S1. In addition, *iCre1355* in SBML format is provided in Data S2.

Flux variability analysis (Mahadevan and Schilling, 2003) was used to determine minimum and maximum allowable flux through each reaction in the network under optimal growth conditions.

Integration of transcriptomics datasets

To make predictions about condition-dependent growth and flux distributions, we used a combination of E-flux (Colijn et al., 2009) and iMAT (Shlomi et al., 2008; Zur et al., 2010). E-flux linearly scales the minimum and maximum allowable flux through reactions in the networks based on transcriptomics data prior to optimizing for the chosen objective function (e.g., growth rate). On the other hand, iMAT uses gene (or protein) expression data to partition reactions in the network into highly and lowly expressed reactions, based on the expression of the genes encoding the catalyzing enzymes, and then maximizes the agreement between the predicted flux distribution and the pre-assigned activity states of the reactions. We implemented E-flux using previously published transcriptomics data for *C. reinhardtii* by linearly scaling the \mathbf{V}_{\min} and \mathbf{V}_{\max} to measured FPKM values for the genes encoding the enzymes associated with the given reaction. To prevent genes with extremely high FPKM counts (outliers) from skewing the flux ranges, the 99th percentile of all FPKMs was taken as the highest count (Max_{fpkm}) (typically ~ 600 FPKM). Thus, for a reaction catalyzed by a single enzyme (E), the upper bound (\mathbf{V}_{\max}) was calculated as:

$$\mathbf{V}_{\max} = \min((E_{\text{fpkm}}/\text{Max}_{\text{fpkm}}) * 100 \text{ gDW}^{-1} \text{ h}^{-1}), 100 \text{ gDW}^{-1} \text{ h}^{-1})$$

For reactions with complex GPR rules, the cumulative FPKM for that reaction was determined based on the following rules: (i) For reactions catalyzed by multiple isozymes (e.g., E1 OR E2 OR E3), the E_{fpkm} was determined using the FPKM of the gene with the

highest count (i.e., $E_{fpkm} = \max(E1_{fpkm}, E2_{fpkm}, E3_{fpkm})$); (ii) For reactions catalyzed by multi-subunit enzymes (e.g., E1 AND E2 AND E3) E_{fpkm} was determined using the FPKM of the gene with the lowest count (i.e., $E_{fpkm} = \min(E1_{fpkm}, E2_{fpkm}, E3_{fpkm})$); (iii) For more complex GPR rules involving a combination of OR and AND relationships (e.g., E1 AND (E2 OR E3)), rule (i) and (ii) were combined (i.e., $E_{fpkm} = \min(E1_{fpkm}, \max(E2_{fpkm}, E3_{fpkm}))$). V_{min} was set to $-V_{max}$ for reversible reactions or $0 \text{ gDW}^{-1} \text{ h}^{-1}$ for irreversible reactions. To enable prediction of growth after imposing the E-flux constraints, four reactions (PSIIred, PSIIblue, ATPSm, ANNA and DHRO) were left unconstrained i.e., V_{max} set to $100 \text{ gDW}^{-1} \text{ h}^{-1}$. This need arose due to the presence of very lowly expressed peripheral subunits that are included in the GPR rules for large multiple subunit complexes such as ATP synthase. A similar set of calculations was conducted for microarray datasets (Mettler et al., 2014), with a few modifications. Firstly, the log transformed intensity values provided in (Mettler et al., 2014) were back transformed to their original values to generate a larger dynamic range. We then selected the 95th percentile of all intensities set as the highest intensity to exclude outliers. Flux predictions were made either by maximizing for biomass (or TAG) production (E-flux) or by maximizing the agreement between gene expression and predicted flux distribution (iMAT) in context of the E-flux bounds (Eflux+iMAT).

To implement iMAT, reactions were divided into high and low expression reactions using thresholds. Reaction(s) catalyzed by an enzyme(s) with an E_{fpkm} less than or equal to 3 FPKM, were assigned as being low expression, while reactions catalyzed by an enzyme(s) with E_{fpkm} greater than or equal to 95th percentile of all FPKMs in the control condition (typically ~75 FPKM) were assigned as high expression. The rules for determining the E_{fpkm} of a reaction were the same as described above. For microarray data, a reaction(s) catalyzed by an enzyme(s) with E_{fpkm} greater than or equal to the ~84th percentile of intensities in the control condition was assigned as high expression, while those lower than the ~16th percentile were assigned as low expression. These thresholds were chosen because RNA-seq datasets are typically very skewed to the right, while microarray datasets are more normally distributed (Oshlack et al., 2010; Guo et al., 2013; Zwiener et al., 2014) (microarray upper/lower bound equal to one standard deviation from mean). An assessment on the impact of the selected thresholds on the predictions using this approach indicated that the predictive accuracy, calculated as the mean squared error (i.e., the mean of the squared difference between predicted and observed growth rates), was relatively robust to the selected high expression threshold (Figure S5a). On the other hand, the selected low expression threshold cutoff can have a significant impact on predictive accuracy, with thresholds greater than or equal to the 20th percentile resulting in a significant increase in the mean squared error (i.e., significant decrease in predictive accuracy) (Figure S6b). iMAT was formulated as previously described and epsilon was set to 0.001 (Shlomi et al., 2008).

The flux capacity of reactions in *iCre1355* was determined using FVA, which was used to determine the effective minimum ($V_{min,eff}$) and maximum ($V_{max,eff}$) allowable flux through each reaction in the network based on constraints throughout the network. The predicted flux capacity through each reaction was then calculated as: $\max(\text{abs}(V_{max,eff}), \text{abs}(V_{min,eff}))$, i.e., the maximum allowable flux, in either direction, permitted through a given reaction within the context of the applied network constraints (including transcriptional constraints

applied using Eflux). To assess the impact of relative changes in reaction flux capacity at a pathway level, we compared the predicted flux capacities for all reactions in a pathway at each time point to 0 m. Using a threshold of 0.01 mmol gDW⁻¹ h⁻¹, the pathway-level change in capacity at each time point was taken as the difference between total number of reactions in a given pathway with significantly increased predicted flux capacity and the total number of reactions in that pathway with significantly reduced flux capacity.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We would like to thank Dr. Julie Bletz for careful reading of the manuscript and for helpful comments. This work was funded in part by DOE-ABY (DEEE0006315) (SI, JV, ALGdL, WC, NSB, NDP), NIH Center for Systems Biology/2P50 GM076547 (NSB, NDP) and the Camille Dreyfus Teacher-Scholar program (NDP). We also acknowledge financial support from the German Ministry for Research and Education within the framework of the GerontoSys initiative (grant FKZ 0315581D) and GlioPATH (grant FKZ 01ZX1402C) (SS).

References

- Alper H, Jin YS, Moxley JF, Stephanopoulos G. Identifying gene targets for the metabolic engineering of lycopene biosynthesis in *Escherichia coli*. *Metab Eng.* 2005; 7:155–164. [PubMed: 15885614]
- Atteia A, Adrait A, Brugiere S, Tardif M, van Lis R, Deusch O, Dagan T, Kuhn L, Gontero B, Martin W, Garin J, Joyard J, Rolland N. A proteomic survey of *Chlamydomonas reinhardtii* mitochondria sheds new light on the metabolic plasticity of the organelle and on the nature of the alpha-proteobacterial mitochondrial ancestor. *Mol Biol Evol.* 2009; 26:1533–1548. [PubMed: 19349646]
- Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A, Marshall M, Moxon S, Sonnhammer EL, Studholme DJ, Yeats C, Eddy SR. The Pfam protein families database. *Nucleic Acids Res.* 2004; 32:D138–141. [PubMed: 14681378]
- Bishop CL, Purton S, Nugent JH. Molecular analysis of the *Chlamydomonas* nuclear gene encoding PsbW and demonstration that PsbW is a subunit of photosystem II, but not photosystem I. *Plant Mol Biol.* 2003; 52:285–289. [PubMed: 12856936]
- Blaby IK, Blaby-Haas CE, Tourasse N, Hom EF, Lopez D, Aksoy M, Grossman A, Umen J, Dutcher S, Porter M, King S, Witman GB, Stanke M, Harris EH, Goodstein D, Grimwood J, Schmutz J, Vallon O, Merchant SS, Prochnik S. The *Chlamydomonas* genome project: a decade on. *Trends Plant Sci.* 2014; 19:672–680. [PubMed: 24950814]
- Blaby IK, Glaesener AG, Mettler T, Fitz-Gibbon ST, Gallaher SD, Liu B, Boyle NR, Kropat J, Stitt M, Johnson S, Benning C, Pellegrini M, Casero D, Merchant SS. Systems-level analysis of nitrogen starvation-induced modifications of carbon metabolism in a *Chlamydomonas reinhardtii* starchless mutant. *Plant Cell.* 2013; 25:4305–4323. [PubMed: 24280389]
- Bolling C, Fiehn O. Metabolite profiling of *Chlamydomonas reinhardtii* under nutrient deprivation. *Plant Physiol.* 2005; 139:1995–2005. [PubMed: 16306140]
- Boyle NR, Morgan JA. Flux balance analysis of primary metabolism in *Chlamydomonas reinhardtii*. *BMC Syst Biol.* 2009; 3:4. [PubMed: 19128495]
- Boyle NR, Page MD, Liu B, Blaby IK, Casero D, Kropat J, Cokus SJ, Hong-Hermesdorf A, Shaw J, Karpowicz SJ, Gallaher SD, Johnson S, Benning C, Pellegrini M, Grossman A, Merchant SS. Three acyltransferases and nitrogen-responsive regulator are implicated in nitrogen starvation-induced triacylglycerol accumulation in *Chlamydomonas*. *J Biol Chem.* 2012; 287:15811–15825. [PubMed: 22403401]
- Branco-Price C, Kaiser KA, Jang CJ, Larive CK, Bailey-Serres J. Selective mRNA translation coordinates energetic and metabolic adjustments to cellular oxygen deprivation and reoxygenation in *Arabidopsis thaliana*. *Plant J.* 2008; 56:743–755. [PubMed: 18665916]

- Cardol P, Gonzalez-Halphen D, Reyes-Prieto A, Baurain D, Matagne RF, Remacle C. The mitochondrial oxidative phosphorylation proteome of *Chlamydomonas reinhardtii* deduced from the Genome Sequencing Project. *Plant Physiol.* 2005; 137:447–459. [PubMed: 15710684]
- Castruita M, Casero D, Karpowicz SJ, Kropat J, Vieler A, Hsieh SI, Yan W, Cokus S, Loo JA, Benning C, Pellegrini M, Merchant SS. Systems biology approach in *Chlamydomonas* reveals connections between copper nutrition and multiple metabolic steps. *Plant Cell.* 2011; 23:1273–1292. [PubMed: 21498682]
- Chang RL, Ghamsari L, Manichaikul A, Hom EF, Balaji S, Fu W, Shen Y, Hao T, Palsson BO, Salehi-Ashtiani K, Papin JA. Metabolic network reconstruction of *Chlamydomonas* offers insight into light-driven algal metabolism. *Mol Syst Biol.* 2011; 7:518. [PubMed: 21811229]
- Chen F, Johns MR. Substrate inhibition of *Chlamydomonas reinhardtii* by acetate in heterotrophic culture. *Process Biochemistry.* 1994; 29:245–252.
- Chisti Y. Biodiesel from microalgae. *Biotechnol Adv.* 2007; 25:294–306. [PubMed: 17350212]
- Colijn C, Brandes A, Zucker J, Lun DS, Weiner B, Farhat MR, Cheng TY, Moody DB, Murray M, Galagan JE. Interpreting expression data with metabolic flux models: predicting *Mycobacterium tuberculosis* mycolic acid production. *PLoS Comput Biol.* 2009; 5:e1000489. [PubMed: 19714220]
- Croft MT, Warren MJ, Smith AG. Algae need their vitamins. *Eukaryot Cell.* 2006; 5:1175–1183. [PubMed: 16896203]
- Davis MC, Fiehn O, Durnford DG. Metabolic acclimation to excess light intensity in *Chlamydomonas reinhardtii*. *Plant Cell Environ.* 2013; 36:1391–1405. [PubMed: 23346954]
- Duanmu D, Casero D, Dent RM, Gallaher S, Yang W, Rockwell NC, Martin SS, Pellegrini M, Niyogi KK, Merchant SS, Grossman AR, Lagarias JC. Retrograde bilin signaling enables *Chlamydomonas* greening and phototrophic survival. *Proc Natl Acad Sci U S A.* 2013; 110:3621–3626. [PubMed: 23345435]
- Emanuelsson O, Brunak S, von Heijne G, Nielsen H. Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protoc.* 2007; 2:953–971. [PubMed: 17446895]
- Fan J, Andre C, Xu C. A chloroplast pathway for the de novo biosynthesis of triacylglycerol in *Chlamydomonas reinhardtii*. *FEBS Lett.* 2011; 585:1985–1991. [PubMed: 21575636]
- Feely RA, Sabine CL, Lee K, Berelson W, Kleypas J, Fabry VJ, Millero FJ. Impact of anthropogenic CO₂ on the CaCO₃ system in the oceans. *Science.* 2004; 305:362–366. [PubMed: 15256664]
- Guo Y, Sheng Q, Li J, Ye F, Samuels DC, Shyr Y. Large scale comparison of gene expression levels by microarrays and RNAseq using TCGA data. *PLoS One.* 2013; 8:e71462. [PubMed: 23977046]
- Harris EH. *Chlamydomonas* as a model organism. *Annu Rev Plant Physiol Plant Mol Biol.* 2001; 52:363–406. [PubMed: 11337403]
- Hemschemeier A, Duner M, Casero D, Merchant SS, Winkler M, Happe T. Hypoxic survival requires a 2-on-2 hemoglobin in a process involving nitric oxide. *Proc Natl Acad Sci U S A.* 2013; 110:10854–10859. [PubMed: 23754374]
- Henry CS, Zinner JF, Cohoon MP, Stevens RL. iBsu1103: a new genome-scale metabolic model of *Bacillus subtilis* based on SEED annotations. *Genome Biol.* 2009; 10:R69. [PubMed: 19555510]
- Honisch B, Ridgwell A, Schmidt DN, Thomas E, Gibbs SJ, Sluijs A, Zeebe R, Kump L, Martindale RC, Greene SE, Kiessling W, Ries J, Zachos JC, Royer DL, Barker S, Marchitto TM Jr, Moyer R, Pelejero C, Ziveri P, Foster GL, Williams B. The geological record of ocean acidification. *Science.* 2012; 335:1058–1063. [PubMed: 22383840]
- Iwai M, Ikeda K, Shimojima M, Ohta H. Enhancement of extraplastidic oil synthesis in *Chlamydomonas reinhardtii* using a type-2 diacylglycerol acyltransferase with a phosphorus starvation-inducible promoter. *Plant Biotechnol J.* 2014; 12:808–819. [PubMed: 24909748]
- Jones PC, Jiang W, Fillingame RH. Arrangement of the multicopy H⁺-translocating subunit c in the membrane sector of the *Escherichia coli* F1F0 ATP synthase. *J Biol Chem.* 1998; 273:17178–17185. [PubMed: 9642286]
- Kanehisa M, Goto S, Kawashima S, Nakaya A. The KEGG databases at GenomeNet. *Nucleic Acids Res.* 2002; 30:42–46. [PubMed: 11752249]

- Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M. Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res.* 2014; 42:D199–205. [PubMed: 24214961]
- Kellis M, Wold B, Snyder MP, Bernstein BE, Kundaje A, Marinov GK, Ward LD, Birney E, Crawford GE, Dekker J, Dunham I, Elnitski LL, Farnham PJ, Feingold EA, Gerstein M, Giddings MC, Gilbert DM, Gingeras TR, Green ED, Guigo R, Hubbard T, Kent J, Lieb JD, Myers RM, Pazin MJ, Ren B, Stamatoyannopoulos JA, Weng Z, White KP, Hardison RC. Defining functional DNA elements in the human genome. *Proc Natl Acad Sci U S A.* 2014; 111:6131–6138. [PubMed: 24753594]
- Kim J, Reed JL. RELATCH: relative optimality in metabolic networks explains robust metabolic and regulatory responses to perturbations. *Genome Biol.* 2012; 13:R78. [PubMed: 23013597]
- Kleessen S, Irgang S, Klie S, Giavalisco P, Nikoloski Z. Integration of transcriptomics and metabolomics data specifies the metabolic response of *Chlamydomonas* to rapamycin treatment. *Plant J.* 2015; 81:822–835. [PubMed: 25600836]
- Kliphuis AM, Klok AJ, Martens DE, Lamers PP, Janssen M, Wijffels RH. Metabolic modeling of *Chlamydomonas reinhardtii*: energy requirements for photoautotrophic growth and maintenance. *J Appl Phycol.* 2011; 24:253–266. [PubMed: 22427720]
- Krasensky J, Jonak C. Drought, salt, and temperature stress-induced metabolic rearrangements and regulatory networks. *J Exp Bot.* 2012; 63:1593–1608. [PubMed: 22291134]
- Lecler R, Vigeolas H, Emonds-Alt B, Cardol P, Remacle C. Characterization of an internal type-II NADH dehydrogenase from *Chlamydomonas reinhardtii* mitochondria. *Curr Genet.* 2012; 58:205–216. [PubMed: 22814755]
- Lewis NS, Nocera DG. Powering the planet: chemical challenges in solar energy utilization. *Proc Natl Acad Sci U S A.* 2006; 103:15729–15735. [PubMed: 17043226]
- Lin H, Kwan AL, Dutcher SK. Synthesizing and salvaging NAD: lessons learned from *Chlamydomonas reinhardtii*. *PLoS Genet.* 2010; 6:e1001105. [PubMed: 20838591]
- Mahadevan R, Schilling CH. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metab Eng.* 2003; 5:264–276. [PubMed: 14642354]
- Marshall WF. Basal bodies platforms for building cilia. *Curr Top Dev Biol.* 2008; 85:1–22. [PubMed: 19147000]
- McCloskey D, Palsson BO, Feist AM. Basic and applied uses of genome-scale metabolic network reconstructions of *Escherichia coli*. *Mol Syst Biol.* 2013; 9:661. [PubMed: 23632383]
- McGlade C, Ekins P. The geographical distribution of fossil fuels unused when limiting global warming to 2 degrees C. *Nature.* 2015; 517:187–190. [PubMed: 25567285]
- Menacho A, Vega JM. Effect of nitrogen starvation on ammonium assimilation by *Chlamydomonas reinhardtii*. *Physiol Plant.* 1989; 75:285–289.
- Merchant SS, Prochnik SE, Vallon O, Harris EH, Karpowicz SJ, Witman GB, Terry A, Salamov A, Fritz-Laylin LK, Marechal-Drouard L, Marshall WF, Qu LH, Nelson DR, Sanderfoot AA, Spalding MH, Kapitonov VV, Ren Q, Ferris P, Lindquist E, Shapiro H, Lucas SM, Grimwood J, Schmutz J, Cardol P, Cerutti H, Chanfreau G, Chen CL, Cognat V, Croft MT, Dent R, Dutcher S, Fernandez E, Fukuzawa H, Gonzalez-Ballester D, Gonzalez-Halphen D, Hallmann A, Hanikenne M, Hippler M, Inwood W, Jabbari K, Kalanon M, Kuras R, Lefebvre PA, Lemaire SD, Lobanov AV, Lohr M, Manuell A, Meier I, Mets L, Mittag M, Mittelmeier T, Moroney JV, Moseley J, Napoli C, Nedelcu AM, Niyogi K, Novoselov SV, Paulsen IT, Pazour G, Purton S, Ral JP, Riano-Pachon DM, Riekhof W, Rymarquis L, Schroda M, Stern D, Umen J, Willows R, Wilson N, Zimmer SL, Allmer J, Balk J, Bisova K, Chen CJ, Elias M, Gendler K, Hauser C, Lamb MR, Ledford H, Long JC, Minagawa J, Page MD, Pan J, Pootakham W, Roje S, Rose A, Stahlberg E, Terauchi AM, Yang P, Ball S, Bowler C, Dieckmann CL, Gladyshev VN, Green P, Jorgensen R, Mayfield S, Mueller-Roeber B, Rajamani S, Sayre RT, Brokstein P, Dubchak I, Goodstein D, Hornick L, Huang YW, Jhaveri J, Luo Y, Martinez D, Ngau WC, Otilar B, Poliakov A, Porter A, Szajkowski L, Werner G, Zhou K, Grigoriev IV, Rokhsar DS, Grossman AR. The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. *Science.* 2007; 318:245–250. [PubMed: 17932292]
- Mettler T, Muhlhaus T, Hemme D, Schottler MA, Rupprecht J, Idoine A, Veyel D, Pal SK, Yaneva-Roder L, Winck FV, Sommer F, Vosloh D, Seiwert B, Erban A, Burgos A, Arvidsson S,

- Schonfelder S, Arnold A, Gunther M, Krause U, Lohse M, Kopka J, Nikoloski Z, Mueller-Roeber B, Willmitzer L, Bock R, Schroda M, Stitt M. Systems analysis of the response of photosynthesis, metabolism, and growth to an increase in irradiance in the photosynthetic model organism *Chlamydomonas reinhardtii*. *Plant Cell*. 2014; 26:2310–2350. [PubMed: 24894045]
- Miller R, Wu G, Deshpande RR, Vieler A, Gartner K, Li X, Moellering ER, Zauner S, Cornish AJ, Liu B, Bullard B, Sears BB, Kuo MH, Hegg EL, Shachar-Hill Y, Shiu SH, Benning C. Changes in transcript abundance in *Chlamydomonas reinhardtii* following nitrogen deprivation predict diversion of metabolism. *Plant Physiol*. 2010; 154:1737–1752. [PubMed: 20935180]
- Milne CB, Kim PJ, Eddy JA, Price ND. Accomplishments in genome-scale in silico modeling for industrial and medical biotechnology. *Biotechnol J*. 2009; 4:1653–1670. [PubMed: 19946878]
- Mittler R. Oxidative stress, antioxidants and stress tolerance. *Trends Plant Sci*. 2002; 7:405–410. [PubMed: 12234732]
- Moulin M, Nguyen GT, Scaife MA, Smith AG, Fitzpatrick TB. Analysis of *Chlamydomonas* thiamin metabolism in vivo reveals riboswitch plasticity. *Proc Natl Acad Sci U S A*. 2013; 110:14622–14627. [PubMed: 23959877]
- Msanne J, Xu D, Konda AR, Casas-Mollano JA, Awada T, Cahoon EB, Cerutti H. Metabolic and gene expression changes triggered by nitrogen deprivation in the photoautotrophically grown microalgae *Chlamydomonas reinhardtii* and *Coccomyxa* sp. C-169. *Phytochemistry*. 2012; 75:50–59. [PubMed: 22226037]
- Muller-Moule P, Havaux M, Niyogi KK. Zeaxanthin deficiency enhances the high light sensitivity of an ascorbate-deficient mutant of *Arabidopsis*. *Plant Physiol*. 2003; 133:748–760. [PubMed: 12972657]
- Nelson, DL.; Cox, MM. *Lehninger: Principles of biochemistry*. W. H. Freeman and Company; New York: 2005.
- Oberhardt MA, Palsson BO, Papin JA. Applications of genome-scale metabolic reconstructions. *Mol Syst Biol*. 2009; 5:320. [PubMed: 19888215]
- Oshlack A, Robinson MD, Young MD. From RNA-seq reads to differential expression results. *Genome Biol*. 2010; 11:220. [PubMed: 21176179]
- Palsson B. The challenges of in silico biology. *Nat Biotechnol*. 2000; 18:1147–1150. [PubMed: 11062431]
- Park JH, Lee KH, Kim TY, Lee SY. Metabolic engineering of *Escherichia coli* for the production of L-valine based on transcriptome analysis and *in silico* gene knockout simulation. *Proc Natl Acad Sci U S A*. 2007; 104:7797–7802. [PubMed: 17463081]
- Park JJ, Wang H, Gargouri M, Deshpande RR, Skepper JN, Holguin FO, Juergens MT, Shachar-Hill Y, Hicks LM, Gang DR. The response of *Chlamydomonas reinhardtii* to nitrogen deprivation: a systems biology analysis. *Plant J*. 2015; 81:611–624. [PubMed: 25515814]
- Pazour GJ, Agrin N, Leszyk J, Witman GB. Proteomic analysis of a eukaryotic cilium. *J Cell Biol*. 2005; 170:103–113. [PubMed: 15998802]
- Perez-Perez ME, Couso I, Crespo JL. Carotenoid deficiency triggers autophagy in the model green alga *Chlamydomonas reinhardtii*. *Autophagy*. 2012; 8:376–388. [PubMed: 22302003]
- Petti AA, Crutchfield CA, Rabinowitz JD, Botstein D. Survival of starving yeast is correlated with oxidative stress response and nonrespiratory mitochondrial function. *Proc Natl Acad Sci U S A*. 2011; 108:E1089–1098. [PubMed: 21734149]
- Pogoryelov D, Reichen C, Klyszejko AL, Brunisholz R, Muller DJ, Dimroth P, Meier T. The oligomeric state of c rings from cyanobacterial F-ATP synthases varies from 13 to 15. *J Bacteriol*. 2007; 189:5895–5902. [PubMed: 17545285]
- Recht L, Topfer N, Batushansky A, Sikron N, Gibon Y, Fait A, Nikoloski Z, Boussiba S, Zarka A. Metabolite profiling and integrative modeling reveal metabolic constraints for carbon partitioning under nitrogen starvation in the green algae *Haematococcus pluvialis*. *J Biol Chem*. 2014; 289:30387–30403. [PubMed: 25183014]
- Reed JL. Shrinking the metabolic solution space using experimental datasets. *PLoS Comput Biol*. 2012; 8:e1002662. [PubMed: 22956899]

- Reijnders MJ, van Heck RG, Lam CM, Scaife MA, dos Santos VA, Smith AG, Schaap PJ. Green genes: bioinformatics and systems-biology innovations drive algal biotechnology. *Trends Biotechnol.* 2014; 32:617–626. [PubMed: 25457388]
- Rochaix JD. *Chlamydomonas*, a model system for studying the assembly and dynamics of photosynthetic complexes. *FEBS Lett.* 2002; 529:34–38. [PubMed: 12354609]
- Sanz-Luque E, Ocana-Calahorra F, Llamas A, Galvan A, Fernandez E. Nitric oxide controls nitrate and ammonium assimilation in *Chlamydomonas reinhardtii*. *J Exp Bot.* 2013; 64:3373–3383. [PubMed: 23918969]
- Sayre R. Microalgae: the potential for carbon capture. *Bioscience.* 2010; 60:722–727.
- Scott SA, Davey MP, Dennis JS, Horst I, Howe CJ, Lea-Smith DJ, Smith AG. Biodiesel from algae: challenges and prospects. *Curr Opin Biotechnol.* 2010; 21:277–286. [PubMed: 20399634]
- Seelert H, Poetsch A, Dencher NA, Engel A, Stahlberg H, Muller DJ. Structural biology. Proton-powered turbine of a plant motor. *Nature.* 2000; 405:418–419. [PubMed: 10839529]
- Shlomi T, Cabili MN, Herrgard MJ, Palsson BO, Ruppin E. Network-based prediction of human tissue-specific metabolism. *Nat Biotechnol.* 2008; 26:1003–1010. [PubMed: 18711341]
- Small I, Peeters N, Legeai F, Lurin C. Predotar: A tool for rapidly screening proteomes for N-terminal targeting sequences. *Proteomics.* 2004; 4:1581–1590. [PubMed: 15174128]
- Spalding MH. Microalgal carbon-dioxide-concentrating mechanisms: *Chlamydomonas* inorganic carbon transporters. *J Exp Bot.* 2008; 59:1463–1473. [PubMed: 17597098]
- Stock D, Gibbons C, Arechaga I, Leslie AG, Walker JE. The rotary mechanism of ATP synthase. *Curr Opin Struct Biol.* 2000; 10:672–679. [PubMed: 11114504]
- Terashima M, Specht M, Hippler M. The chloroplast proteome: a survey from the *Chlamydomonas reinhardtii* perspective with a focus on distinctive features. *Curr Genet.* 2011; 57:151–168. [PubMed: 21533645]
- Terashima M, Specht M, Naumann B, Hippler M. Characterizing the anaerobic response of *Chlamydomonas reinhardtii* by quantitative proteomics. *Mol Cell Proteomics.* 2010; 9:1514–1532. [PubMed: 20190198]
- Theiss C, Bohley P, Voigt J. Regulation by polyamines of ornithine decarboxylase activity and cell division in the unicellular green alga *Chlamydomonas reinhardtii*. *Plant Physiol.* 2002; 128:1470–1479. [PubMed: 11950995]
- Thiele I, Palsson BO. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat Protoc.* 2010; 5:93–121. [PubMed: 20057383]
- Urzica EI, Adler LN, Page MD, Linster CL, Arbing MA, Casero D, Pellegrini M, Merchant SS, Clarke SG. Impact of oxidative stress on ascorbate biosynthesis in *Chlamydomonas* via regulation of the VTC2 gene encoding a GDP-L-galactose phosphorylase. *J Biol Chem.* 2012; 287:14234–14245. [PubMed: 22393048]
- Urzica EI, Casero D, Yamasaki H, Hsieh SI, Adler LN, Karpowicz SJ, Blaby-Haas CE, Clarke SG, Loo JA, Pellegrini M, Merchant SS. Systems and trans-system level analysis identifies conserved iron deficiency responses in the plant lineage. *Plant Cell.* 2012; 24:3921–3948. [PubMed: 23043051]
- Urzica EI, Vieler A, Hong-Hermesdorf A, Page MD, Casero D, Gallaher SD, Kropat J, Pellegrini M, Benning C, Merchant SS. Remodeling of membrane lipids in iron-starved *Chlamydomonas*. *J Biol Chem.* 2013; 288:30246–30258. [PubMed: 23983122]
- van Lis R, Atteia A, Mendoza-Hernandez G, Gonzalez-Halphen D. Identification of novel mitochondrial protein components of *Chlamydomonas reinhardtii*. A proteomic approach. *Plant Physiol.* 2003; 132:318–330. [PubMed: 12746537]
- Varma A, Palsson BO. Metabolic flux balancing: basic concepts, scientific and practical use. *Nature Biotechnology.* 1994; 12:994–998.
- Wijffels RH, Barbosa MJ. An outlook on microalgal biofuels. *Science.* 2010; 329:796–799. [PubMed: 20705853]
- Zur H, Ruppin E, Shlomi T. iMAT: an integrative metabolic analysis tool. *Bioinformatics.* 2010; 26:3140–3142. [PubMed: 21081510]
- Zwiener I, Frisch B, Binder H. Transforming RNA-Seq data to improve the performance of prognostic gene signatures. *PLoS One.* 2014; 9:e85150. [PubMed: 24416353]

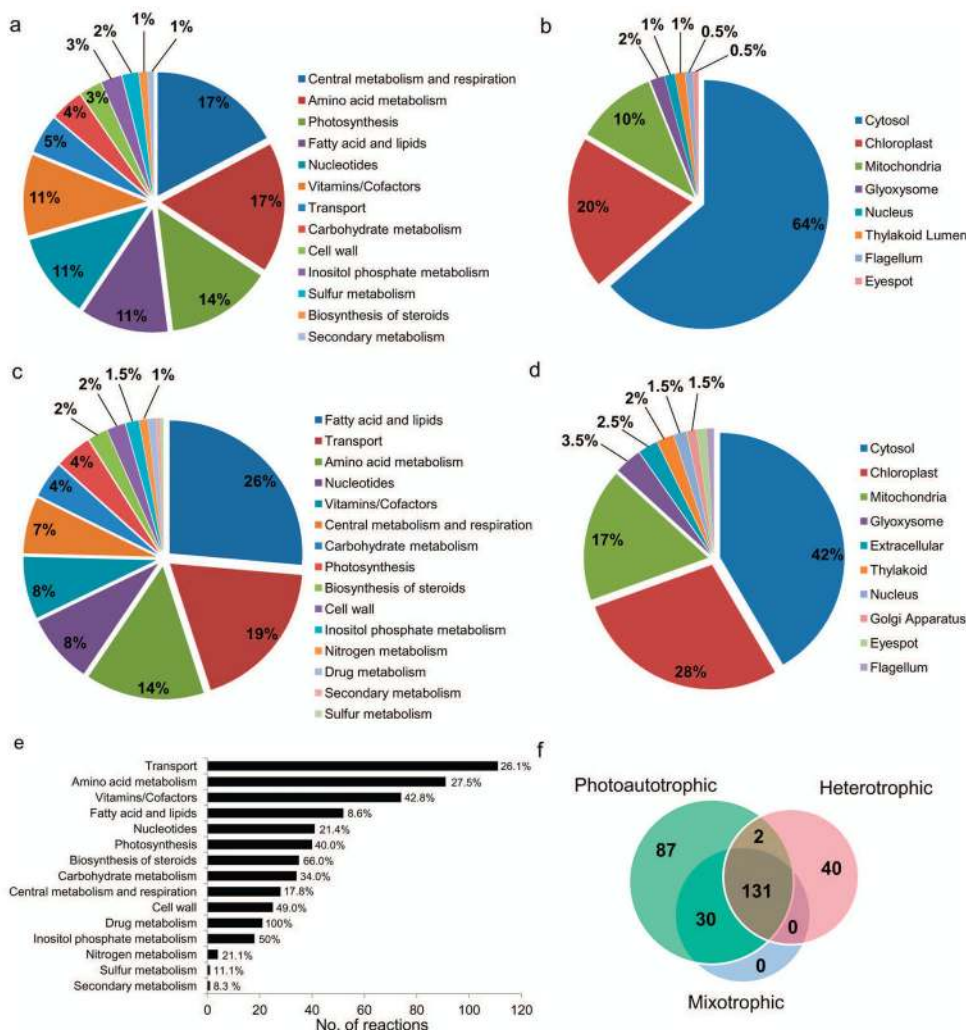


Figure 1. Distribution of genes and reactions in iCre1355

(A) Subsystem distribution of 312 newly identified metabolic genes incorporated in iCre1355. (B) Distribution of the subcellular localization of reactions associated with the 312 newly identified metabolic genes. (C) Subsystem distribution of all reactions in iCre1355. (D) Subcellular localization of all reaction in iCre1355. (E) Subsystem distribution of blocked reactions in iCre1355. Blocked reactions were identified using flux variability analysis with all exchange reactions made reversible. (F) Venn diagram depicting the overlap in predicted essential genes in iCre1355 under photoautotrophic, mixotrophic and heterotrophic conditions.

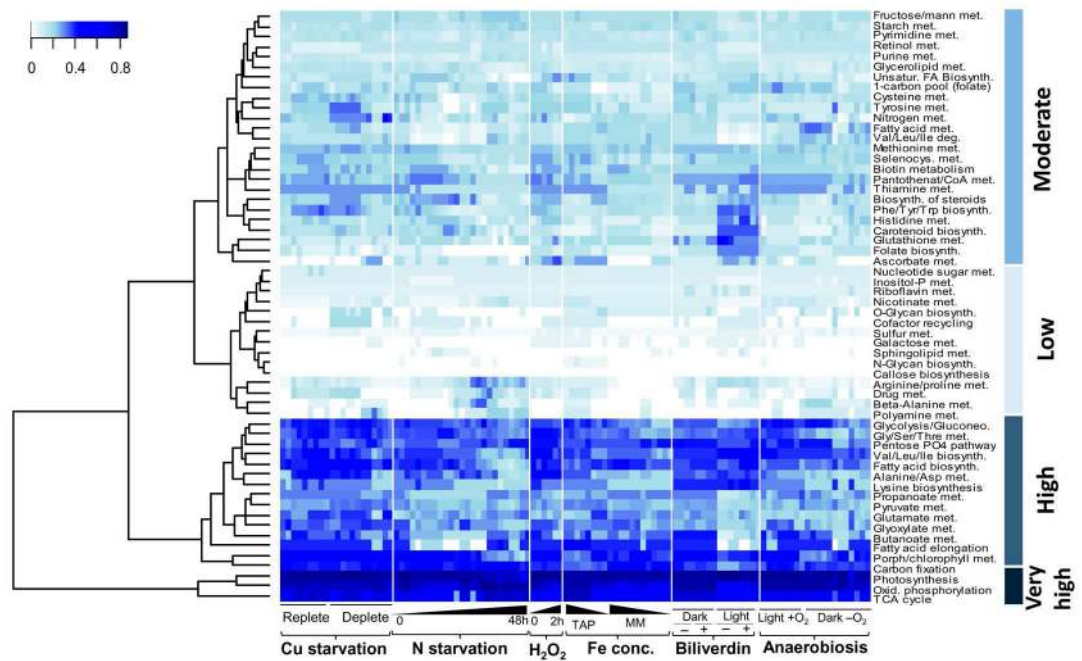


Figure 2. Pathway-level expression of genes in *iCre1355* across conditions

Heatmap depicting the proportion of genes within each pathway in *iCre1355* that are highly expressed across 6 groups of experiments. Pathways were grouped into very highly, highly, moderately and lowly expressed categories based on hierarchical clustering.

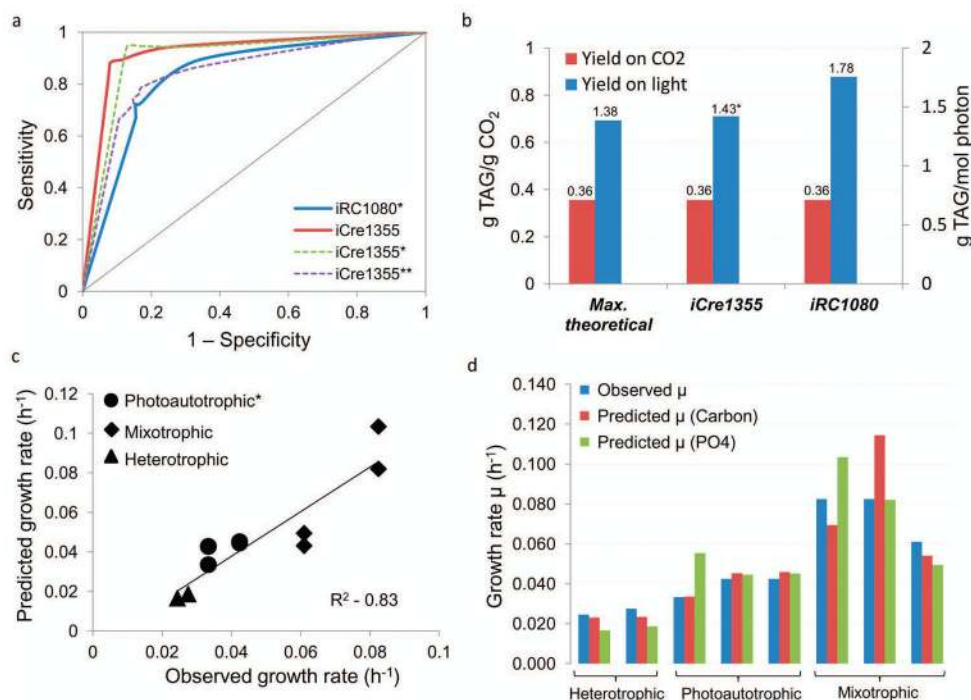


Figure 3. Assessing the performance of *iCre1355*

(A) ROC curves comparing the predictions of *iCre1355* to *iRC1080* for gene deletion phenotype data. * Only 48 out of 81 metabolic mutants could be assessed with *iRC1080*. *iCre1355* was also evaluated with this set of 48 mutants. ** *iCre1355* predictions using biomass reactions from *iRC1080*. (B) Assessment of predicted maximal TAG yield on CO₂ and photon uptake by *iCre1355* and *iRC1080*. * Yields range from 1.33 to 1.43 depending on the number of c subunits (14 to 12 respectively) assumed for ATP synthase. A value of 12 used for this simulation. (C) Comparison of *iCre1355* predicted growth rate to observed growth rates from chemostat grown cells. Measured uptake rates for carbon (CO₂ or Acetate), nitrogen (NH₄) and phosphorus (PO₄) were used as constraints. (D) Predicted growth rates when only CO₂ or PO₄ used as initial constraints.

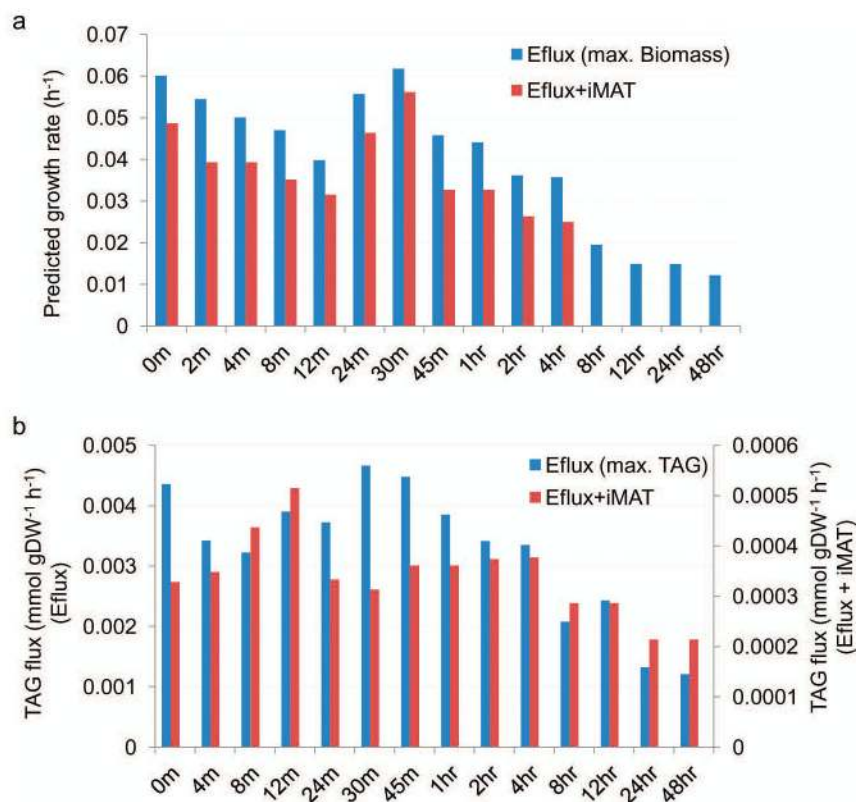


Figure 4. Predicting growth rate and TAG flux during nitrogen starvation

Predicted growth rate (A) and TAG flux (B) during the time course of nitrogen starvation. Predictions were made, with transcriptional data used to set flux capacity bounds, either by maximizing for biomass (E-flux) or maximizing the agreement between gene expression and predicted flux (E-flux+iMAT). The 2 minute time point was an outlier in the TAG flux analysis and thus was omitted to permit better visualization of the other time points.

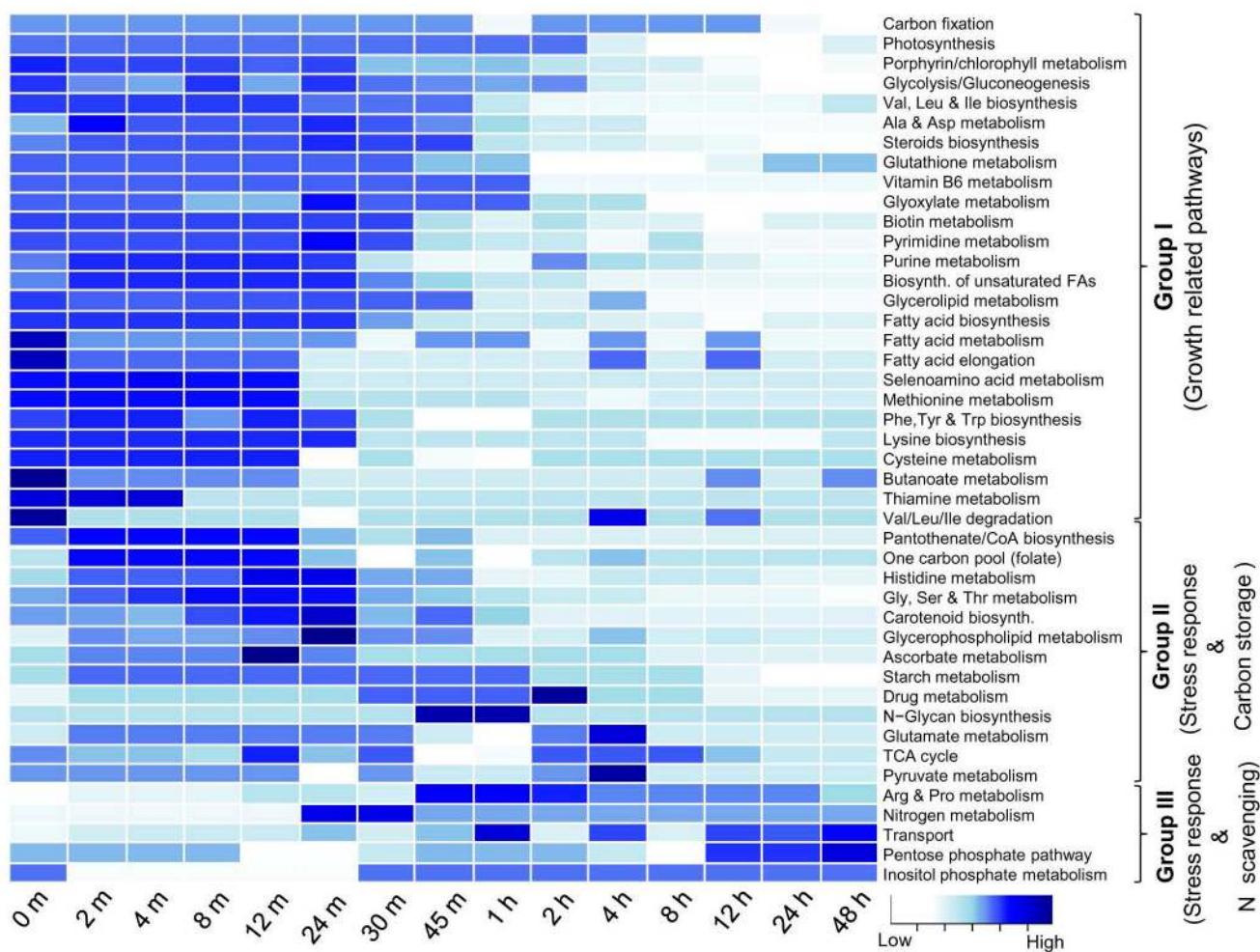


Figure 5. Impact of nitrogen starvation on pathway activity in *iCre1355*
Heat map depicting predicted changes in pathway activity across time during nitrogen starvation. Darker shades indicate greater predicted pathway activity based on an increased number of "high" activity state reactions.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

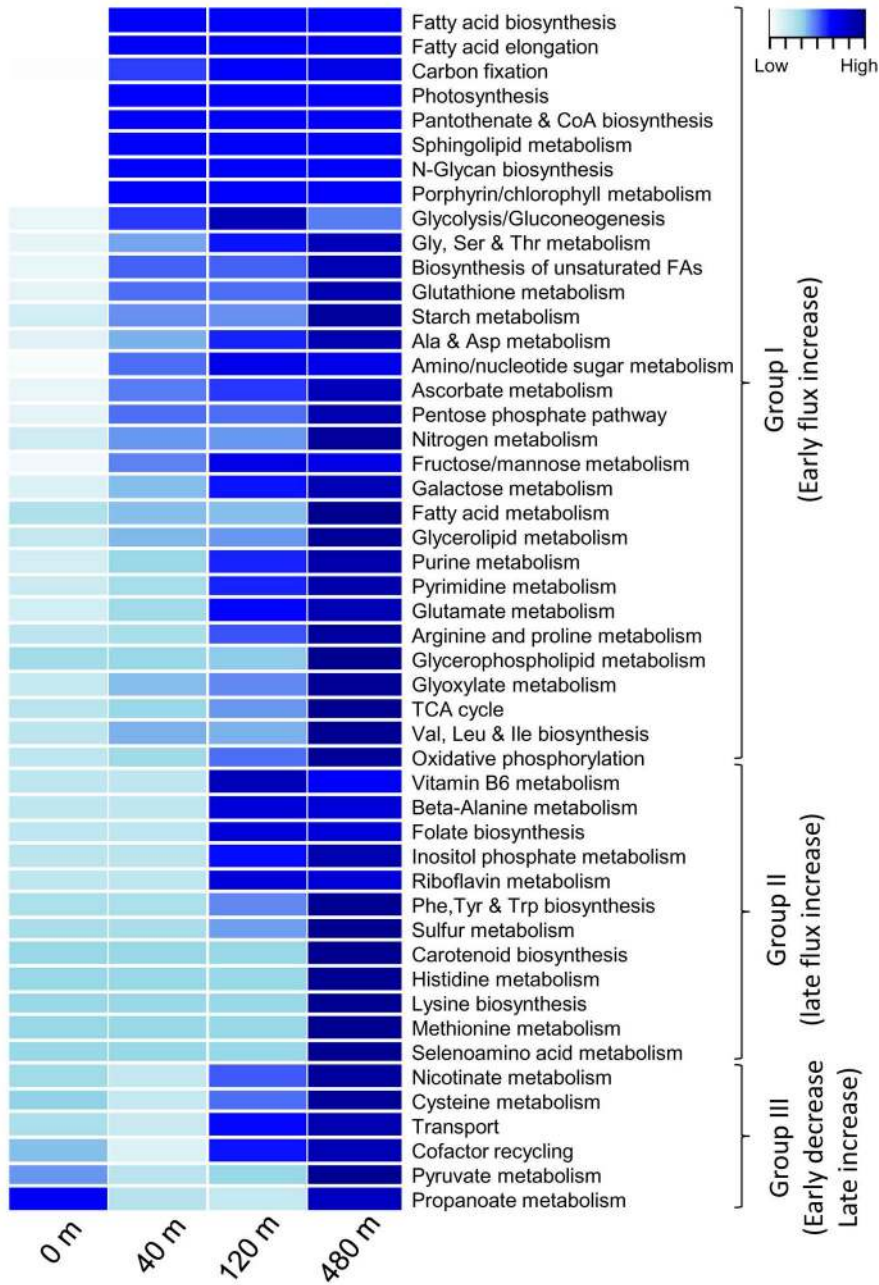


Figure 6. Pathway-level impact of change in light intensity on flux capacity in *iCre1355*
Heatmap depicting predicted changes in pathway activity during the shift from low to a higher light intensity. Pathway-level changes were predicted based on predicted changes in flux capacities of constituent reactions relative to time 0 m.

Table 1Comparison of components in *iCre1355* and *iRC1080*

Components	<i>iCre1355</i>	<i>iRC1080</i>
Genes	1,355	1,073
Alternative transcripts	105	13
Nuclear genes	1,314	1,073
Chloroplast genes	34	0
Mitochondrial genes	7	0
Genome version	JGI version 5.5	JGI version 4
Reactions	2,394	2,190
Enzymatic	1,862	1,717
Experimental evidence-based	779	726
No gene assignment (enzymatic)	114	116
Transport	426	388
Exchange	55	45
Demand	36	26
Reversible	713	656
Irreversible	1,681	1,534
Blocked reactions*	595	613
Compartments	10	10
Metabolites	1,845	1,706
Unique metabolites	1,133	1,068
Cytosol	730	675
Chloroplast	478	457
Mitochondria	308	265

* Reactions that do not carry any flux from flux variability analysis when all exchange reactions made reversible.

Table 2

Uptake rates measured from *C. reinhardtii* cells grown in continuous culture

Conditions	μ (h ⁻¹)	Uptake rates (mmol gDW ⁻¹ h ⁻¹)				
		DW (g/L)*	DIC	NH ₄	PO ₄	Acetate
Autotrophic_Rep1	0.033	0.115	1.289	0.569	0.032	NA
Autotrophic_Rep2	0.033	0.080	ND	0.734	0.024	NA
Autotrophic_Rep3	0.042	0.069	1.752	0.514	0.025	NA
Autotrophic_Rep4	0.042	0.060	1.775	0.486	0.026	NA
Mixotrophic_Rep1	0.082	0.048	ND	3.035	0.099	1.364
Mixotrophic_Rep2	0.082	0.078	ND	0.748	0.078	2.261
Mixotrophic_Rep3	0.061	0.100	ND	1.259	0.047	1.059
Mixotrophic_Rep4	0.061	0.166	ND	1.071	0.041	1.234
Heterotrophic_Rep1	0.024	0.137	ND	0.439	0.010	1.579
Heterotrophic_Rep2	0.028	0.137	ND	0.454	0.012	1.599

* DW – Dry weight adjusted for 5.68% ash content (Klipphuis et al., 2011).

ND - Not determined (macronutrient not quantified). NA - Not applicable (macronutrient not component of media).