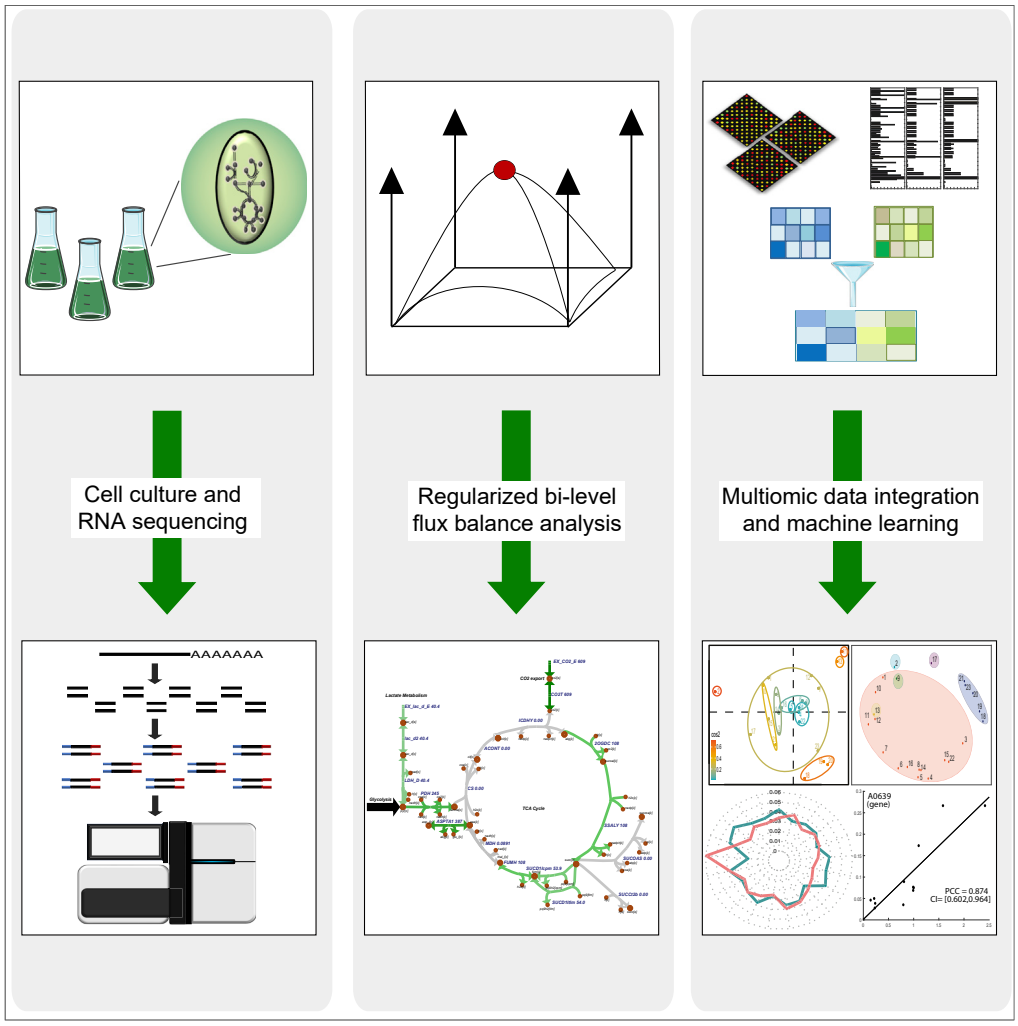
iScience

CellPress
OPEN ACCESS

Article

A Hybrid Flux Balance Analysis and Machine Learning Pipeline Elucidates Metabolic Adaptation in Cyanobacteria



Supreeta Vijayakumar, Pattanathu K.S.M. Rahman, Claudio Angione

c.angione@tees.ac.uk

HIGHLIGHTS

A pipeline for metabolic modeling in *Synechococcus* sp. PCC 7002 is presented

Metabolic fluxes display clear differences in pathway activity across conditions

Omic-informed GSMMs provide critical mechanistic details within machine learning

Combining GSMM and machine learning improves methods based on transcriptomics alone

Vijayakumar et al., iScience *23*, 101818
December 18, 2020
https://doi.org/10.1016/j.isci.2020.101818

iScience

CellPress
OPEN ACCESS

Article

A Hybrid Flux Balance Analysis and Machine Learning Pipeline Elucidates Metabolic Adaptation in Cyanobacteria

Supreeta Vijayakumar,[1] Pattanathu K.S.M. Rahman,[2,3] and Claudio Angione[1,4,5,6,*]

SUMMARY

Machine learning has recently emerged as a promising tool for inferring multi-omic relationships in biological systems. At the same time, genome-scale metabolic models (GSMMs) can be integrated with such multi-omic data to refine phenotypic predictions. In this work, we use a multi-omic machine learning pipeline to analyze a GSMM of *Synechococcus* sp. PCC 7002, a cyanobacterium with large potential to produce renewable biofuels. We use regularized flux balance analysis to observe flux response between conditions across photosynthesis and energy metabolism. We then incorporate principal-component analysis, *k*-means clustering, and LASSO regularization to reduce dimensionality and extract key cross-omic features. Our results suggest that combining metabolic modeling with machine learning elucidates mechanisms used by cyanobacteria to cope with fluctuations in light intensity and salinity that cannot be detected using transcriptomics alone. Furthermore, GSMMs introduce critical mechanistic details that improve the performance of omic-based machine learning methods.

INTRODUCTION

In the field of systems biology, several approaches have been proposed to capture the enormous complexity of biological systems by utilizing mathematical modeling and computational methods, with the goal of amalgamating the information required to build and refine predictive models. The challenges presented by such an undertaking are numerous and persistent owing to the size, format, scale, and variation of the disparate data types. Among these, metabolism is currently the only biological layer that can be modeled genome-wide (O'Brien et al., 2015; Haas et al., 2017). Constraint-based reconstruction and analysis methods are commonly used to express metabolic flux through biochemical reactions based on knowledge of reaction stoichiometry. Flux balance analysis (FBA) is particularly suitable for modeling metabolic networks at the genome scale, as the definition of kinetic parameters and metabolite concentrations is not a key requisite.

In recent years, genome-scale metabolic models (GSMMs) have been integrated with multiple data types, including omics, codon usage, enzyme costs, and limited resource availability (Abedpour and Kollmann, 2015; Opdam et al., 2017; Kashaf et al., 2017; Wortel et al., 2018; Tian and Reed, 2018; Angione, 2019). This serves to exploit the large volume of experimental data being generated from high-throughput omics technologies. In doing so, additional constraints can be applied during FBA to shrink the solution space (Reed, 2012), thus providing a more accurate representation of metabolic capability as a greater number of factors can be considered to explain cellular behavior. This can prove useful in refining phenotypic predictions across various environmental conditions (Vijayakumar et al., 2017; Sánchez et al., 2017; van der Ark et al., 2017; Angione, 2018) and can predict steps to engineer an organism in a way that optimizes the production of certain metabolites, which is highly applicable in many fields of industrial biotechnology including the production of biofuels, biosurfactants, and pharmaceuticals (Angione et al., 2015; Dougherty et al., 2017; Huang et al., 2017; Fatma et al., 2018; Occhipinti et al., 2018).

Modeling and Metabolic Engineering in Cyanobacteria

Cyanobacteria is a phylum of oxygenic, phototrophic microalgae that need to adapt to constant fluctuations in temperature, salinity, light intensity (or irradiance), and nutrient availability, among other factors

[1]Department of Computer Science and Information Systems, Teesside University, Middlesbrough, North Yorkshire TS1 3BX, UK

[2]Centre for Enzyme Innovation, Institute of Biological and Biomedical Sciences, School of Biological Sciences, University of Portsmouth, Portsmouth, Hampshire PO1 2UP, UK

[3]Tara Biologics, Woking, Surrey GU21 6BP, UK

[4]Centre for Digital Innovation, Teesside University, Middlesbrough TS1 3BX, UK

[5]Healthcare Innovation Centre, Teesside University, Middlesbrough TS1 3BX, UK

[6]Lead Contact

*Correspondence: c.angione@tees.ac.uk

https://doi.org/10.1016/j.isci.2020.101818

(Montgomery, 2017; Blanco-Ameijeiras et al., 2018; Gunde-Cimerman et al., 2018). Metabolic engineering is helping to develop cyanobacteria into photoautotrophic biofactories that can act as production hosts (chassis) for alcohols, carbohydrates, organic acids, fatty acid derivatives, isoprenoids, and many other chemicals (Noreña-Caro and Benton, 2018). However, as such approaches are generally designed with heterotrophic organisms in mind, the metabolic features unique to photoautotrophs must be considered, e.g., pathways relating to photosynthesis and CO_2 fixation (Carroll et al., 2018).

Synechococcus sp. PCC 7002 is a fast-growing cyanobacterium that flourishes in both freshwater and marine environments, owing to its ability to tolerate high light intensity and a wide range of salinities. Harnessing the properties of cyanobacteria has become an important goal in recent years owing to their potential to serve as biocatalysts for the production of renewable biofuels (Hendry et al., 2016). Metabolic modeling of two cyanobacteria, *Arthrospira* and *Synechocystis*, has successfully characterized the use of photosynthetic electron transport components in different light conditions (Toyoshima et al., 2020).

In an industrial setting, *Synechococcus* sp. PCC 7002 has been recommended as the ideal chassis for the mass cultivation of microalgae for biotechnological applications owing to its ease of genetic manipulation as well as its tolerance for high salinity, light intensity, and temperature (Pade and Hagemann, 2014; Clark et al., 2018). These are highly desirable traits in microalgae as they enable cultures to maintain a rapid growth rate in open raceway ponds as well as in photobioreactors, which operate at high temperatures (Ruffing et al., 2016). Within the *Synechococcus* genus, a comparative analysis of slow- and fast-growing strains in terms of their active reactions under phototrophic conditions has been proposed to better inform their development into production hosts (i.e., strain optimization), primarily through maximizing their growth rates (Hendry et al., 2019). In a recent study, Song et al. (2015) completed an integrative analysis of metabolic and gene co-expression networks in *Synechococcus* sp. PCC 7002 by integrating expression data from either continuous cultures or existing studies into a GSMM and deriving fluxes using E-Fmin flux minimization (Song et al., 2014) and MOMA (Segre et al., 2002). Further studies have examined temporal variations in response to varying light intensity and associated conditional dependencies (Rügen et al., 2015; Reimers et al., 2016). These need to be accounted for as constraints in GSMMs designed to simulate the phototrophic growth in cyanobacteria over diurnal cycles and tackle issues associated with resource allocation (Vijayakumar and Angione, 2017).

Genome-scale isotopic non-stationary metabolic flux analysis (INST-MFA) has been utilized to estimate internal metabolic fluxes more accurately in *Synechococcus elongatus* UTEX 2973, toward the aim of establishing factors affecting phototrophic metabolism under optimal growth conditions (Hendry et al., 2019). Similarly, MOMA and INST-13C MFA were used to establish carbon partitioning at intracellular branching points in the central metabolism of a glycogen-deficient *Synechococcus* sp. PCC 7002 mutant (Hendry et al., 2017). Such models benefit significantly from constraints designed using experimentally measured uptake or growth rates for the identification of alternative reactions responsible for the synthesis of metabolites and differences in pathway recruitment and utilization (e.g., for carbon conversion to biomass).

The current state of strain-specific metabolic modeling in cyanobacteria, and the potential of fluxomic data and metabolic engineering, have been recently discussed elsewhere (Angermayr et al., 2015; Oliver et al., 2016; Hendry et al., 2020; Luan et al., 2020; Babele and Young, 2020; Mukherjee et al., 2020; Hitchcock et al., 2020). Within the *Synechococcus* genus, a comparative analysis of slow- and fast-growing strains in terms of their active reactions under phototrophic conditions has been proposed to better inform their development into production hosts (i.e., strain optimization), primarily by maximizing their growth rates (Hendry et al., 2019). A number of novel, non-model strains of *Synechococcus* that have been developed include *Synechococcus* UTEX 2973 (Yu et al., 2015), PCC 11801 (Jaiswal et al., 2018), PCC 11802 (Damini et al., 2020), PCC 11901 (Włodarczyk et al., 2020), and BDU 130192 (Ahmad et al., 2020).

Multi-omic Data Integration in Microalgae

In recent years, synthetic biology has facilitated the modeling of biological processes for genetic engineering. Using synthetic biology tools, algal strains have been designed according to highly specific environmental conditions and yield requirements. Synthetic biologists have been successful in assembling genetic material and manipulating the lipid content of microalgae, as well as maximizing biomass accumulation and biofuel yield (Jagadevan et al., 2018). These results are promising for the biofuels industry from the microalgal perspective (Randhawa et al., 2017). In the context of microalgae, the alteration of lipid

biosynthesis pathways through the induction of a stress response to a change in environment (such as temperature, nutrient limitation, salinity) is a common practice to enhance the production of target compounds, including those that are used to produce workable biofuels (Rawat et al., 2013).

Omics approaches have made a significant contribution to the understanding of the molecular processes of microalgae. Furthermore, the discoveries that omics studies have made, e.g., the identification of genes involved in specific processes, may be vital to the engineering of enhanced microalgae. Through the understanding of transcription levels and gene activation data gathered from transcriptomics, the effectiveness of genetic alterations can be measured as previously achieved for other organisms, allowing for optimization of the target product. For example, if the new gene insert is operating at its optimum, the transcriptomic data should show an increase in the mRNA of the target gene when compared with the wild type (Randhawa et al., 2017). Based on genomic and transcriptomic data, Wang et al. (2019) recently identified a series of neutral sites on the chromosome of *Synechococcus* sp. PCC 7002 for the introduction of novel heterologous genes or pathways without disruption.

Omics techniques can also provide valuable insights into alterations of lipid synthesis pathways that occur as a result of stress conditions in microalgae. Metabolomics studies assess the low-molecular-weight metabolic end products and are highly indicative of response to stresses. Previously, global transcriptomic, proteomic, and metabolomic analyses have aided in identifying adaptations for cyanobacterial salt tolerance in *Synechocystis* sp. PCC 6803 (Pandhal et al., 2009; Wang et al., 2016). An omic-combination approach would allow for optimization of algal engineering, as the data gathered from transcriptomics should show an increase in transcription in the gene of interest that coincides with a reduction in metabolism caused by stress (such as nutrient limitation) highlighted by metabolomics, if the expression of the gene of interest is linked to a metabolic process. The application of omic studies can not only ascertain the effectiveness of any genetic modification but also be used to optimize the scale-up process. With the use of spatial and temporal omics studies of systems such as raceways used for algal growth, a deeper understanding of how algae will perform in various areas of the raceway can be gained, allowing for process optimization (Randhawa et al., 2017).

Aims and Objectives

In this work, we present a pipeline combining metabolic modeling with statistical and machine learning tools (Figure 1) for analyzing a GSMM of the cyanobacterium *Synechococcus* sp. PCC 7002. We characterize *Synechococcus* adaptation mechanisms using an updated GSMM of *iSyp702* containing 728 genes (Hendry et al., 2016), implementing multi-objective FBA with quadratic regularization. We then apply machine learning techniques to identify functionally important genes and reactions. These include PCA, *k*-means clustering, and LASSO regression, which serve not only to identify biologically significant gene transcripts and fluxes but also to relate these features more effectively to growth-promoting or growth-limiting conditions provided by the initial expression data.

Our goal is to show whether, in a predictive setting, features derived from the metabolic model can add meaningful information to the features derived from the transcriptomic data (Zampieri et al., 2019). Therefore, for each method, we will consider the predictions yielded using three sets of features: (1) gene expression only, (2) fluxes only, and (3) gene expression and fluxes combined.

Through LASSO regression, we find that using flux rates to predict growth rates is more effective than using gene transcript values alone. This suggests that GSMMs provide critical details in terms of stoichiometry, and the involvement of genes in reactions determines the rate of cellular phototrophic growth as well as other modes of energy utilization (e.g., heterotrophy, mixotrophy) in various environmental conditions.

RESULTS

As highlighted above, our goal is to reconnect metabolism to growth and other cellular objectives using a data-driven multi-view approach that yields biologically reasonable predictions. The results of PCA and *k*-means clustering for flux data are included in Figure 2, whereas results of these analyses for gene transcript data in isolation and gene transcript data combined with fluxes are detailed in Figure 3. Additionally, the results of the pathway-wide analysis of principal components are provided in Figure 4. The highest positive/negative Pearson correlation coefficient (PCC) values for transcript- and flux-only datasets are given

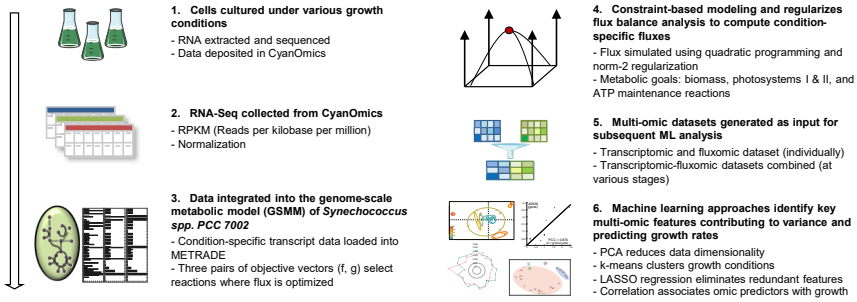


Figure 1. A Multi-omic Machine Learning Pipeline for Prediction and Classification of *Synechococcus* Metabolic Features

(1) RNA sequencing data obtained from *Synechococcus* sp. PCC 7002 cells grown under 23 growth conditions (Ludwig and Bryant, 2011, 2012a,b). (2) Data downloaded from CyanOmics (Yang et al., 2015). (3) Starting from a model recently published by Hendry et al. (2016), the condition-specific GSMMs of *Synechococcus* sp. PCC 7002 are generated by integrating omics data, and three pairs of objectives are optimized for each condition-specific model. (4) Bilevel regularized FBA is conducted using quadratic programming to compute regularized flux distributions. (5) Transcriptomic, fluxomic, and multi-omic (combination) datasets are preprocessed for the machine learning analysis. (6) PCA, *k*-means clustering, LASSO regression, and correlation analysis are applied to identify latent cross-omic patterns in the metabolic adaptation mechanisms. These techniques are applied and compared across three sets of omic features: gene transcripts, condition-specific flux rates, and a combination of both omics.

in Figure 5, whereas mean absolute PCCs for metabolic subsystems or pathways are shown in Figure 6C with the number of reactions in each subsystem specified in Figure 6D. A list of all nonzero LASSO coefficients and the top 10 positive/negative correlation coefficients are given in the Supplemental Information, with the full calculation of these coefficients provided in Data S2. An interpretation of the results for each technique used is provided below.

Regularized Flux Balance Analysis

Compared with transcriptomics, metabolic flux data modeled at the genome-scale provide a more comprehensive, condition-specific view of the phenotype. Therefore, we mapped each RNA sequencing profile measured in 24 growth conditions to a *Synechococcus* GSMM, and we employed a regularized FBA to obtain condition-specific flux distributions (see Transparent Methods in Supplemental Information). To calculate the flux rates more accurately for each condition, several lower and upper bounds were adjusted before performing FBA, according to specific growth media and other requirements described for each growth condition (Ludwig and Bryant, 2011, 2012a, 2012b). The full details of these growth conditions (including composition of growth media, optical density at the time of cell harvestation, mode of energy utilization, availability of oxygen/carbon dioxide, light intensity, salinity, and temperature) are listed in Table S1. The full specification of constraints for each growth condition is given in Data S3.

From the transcriptomic studies listed in Table S1, there were a number of genes that were not transcribed in the control condition but were transcribed specifically under perturbed conditions. Many of these genes have yet to be assigned a particular functional category or encode hypothetical proteins, but many more have been linked to specific pathways and compounds and some have been associated with the adaptation of *Synechococcus* sp. PCC 7002 to atypical environmental or growth conditions.

As shown in Figure 2, apart from the standard control, the highest fluxes through the ATP maintenance reaction (when ATP maintenance is set as the secondary objective) were among conditions that limit growth, such as phosphate limitation, 30°C, and oxidative stress. In dark anoxic, low-salinity, heat shock, phosphate limitation, and mixotrophic conditions, there was no flux for the biomass reaction. However, for all the objective pairs, the flux through the biomass during high light intensity and OD 0.4 (optical density) was higher than the control condition (0.19 mmol gDW^{-1} h^{-1} and 0.093 mmol gDW^{-1} h^{-1} respectively, compared with 0.053 mmol gDW^{-1} h^{-1}). The biomass is likely to be higher at OD 0.4 than OD 0.7 due to adjustment of the photon constraint, which allowed for more transmission of light at lower OD. Apart from the dark anoxic and low O_2 conditions, all fluxes through photosystem II were negligible, but the

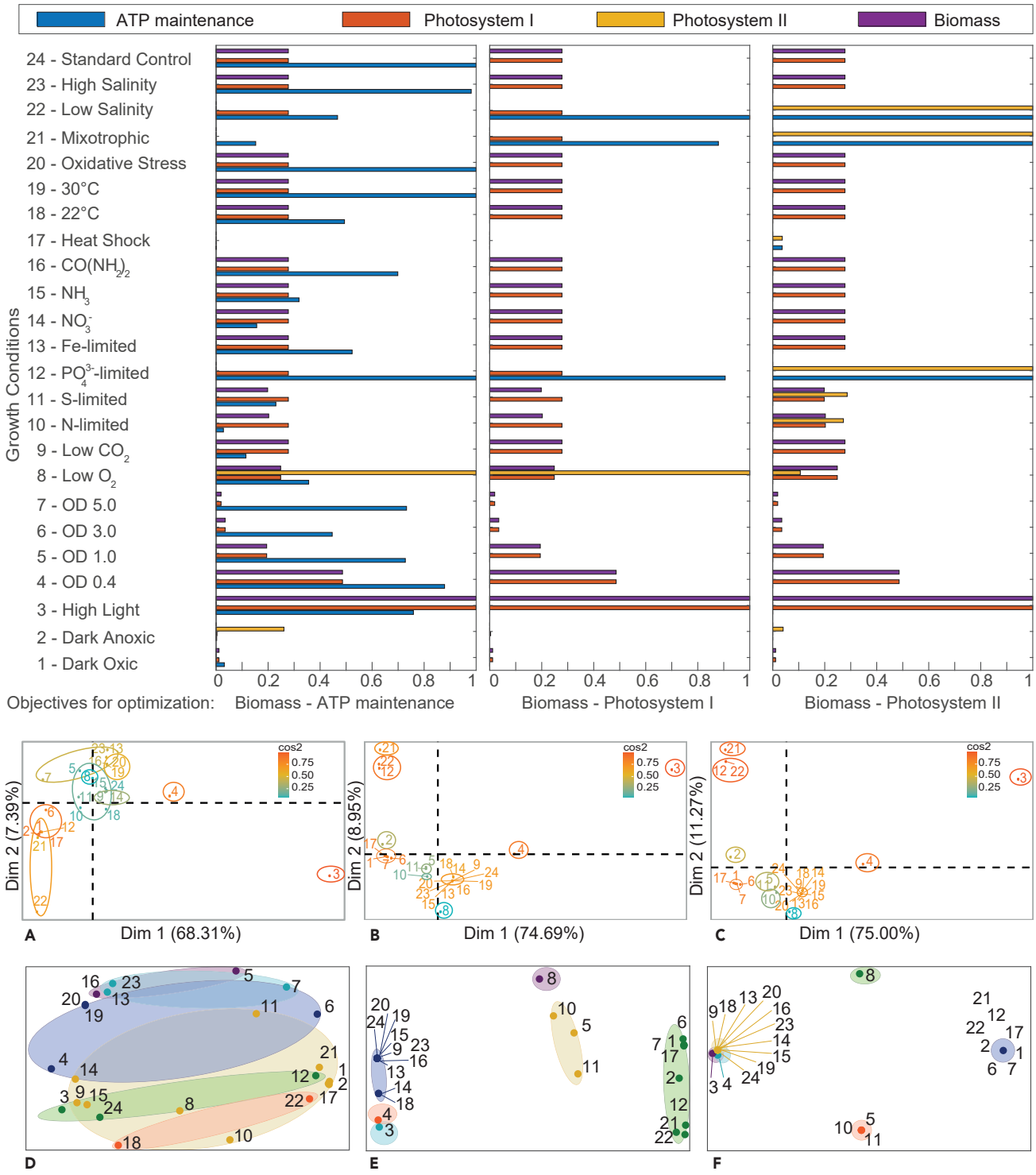


Figure 2. Flux Distributions with Fluxomic PCA and *k*-means

(Top): Flux distributions in the 24 growth conditions considered in this study. Flux distributions for four key reactions: ATP maintenance, photosystem I, photosystem II, and biomass when running FBA using three different pairs of objectives (indicated at the bottom of each plot). Conditions 1–24 correspond to those detailed in Table S1. To better visualize the differences in flux between conditions, flux values were normalized by dividing by the maximal flux (i.e., the flux value for the control condition) for that reaction across all conditions. The full list of flux rates is reported in the Supplemental Information. Regularized FBA correctly predicts reduced growth in sub-optimal conditions and the highest biomass flux is given by the high light condition. (Bottom): Fluxomic principal-component analysis (PCA) and *k*-means. PCA (A–C) and *k*-means clustering (panels D–F) were conducted using the entire flux distribution

Figure 2. *Continued*
(742 reactions). (A) and (D) are associated with Biomass-ATP maintenance fluxes, (B) and (E) with Biomass-Photosystem I fluxes, and (C) and (F) with Biomass-Photosystem II fluxes. For PCA plots, growth conditions are colored according to their cos2 value, which indicates the contribution of the first two components to the squared distance of each condition to the origin (Abdi and Williams, 2010). The higher the cos2 value, the greater the proportion of contribution to the total distance, meaning that the importance of the principal components is greater for that condition. For *k*-means, data are clustered by condition (where the colors of ellipses represent different clusters) and the number of clusters (k=6) was selected following silhouette analysis. Due to co-location of conditions in the two-dimensional plot, not all overlapping points are visible, but the cluster associated with each condition is labeled. The full list of growth conditions and their respective *k*-means clusters are reported in the Supplemental Information.

fluxes through photosystem I were still maintained (0.058 mmol gDW^{-1} h^{-1} in the high light intensity condition as opposed to 0.016 mmol gDW^{-1} h^{-1} in the control).

When photosystem I was set as the secondary objective, a low amount of flux through ATP maintenance reaction was retained in phosphate-limited, heat shock, and low salinity conditions (approximately 0.0002 mmol gDW^{-1} h^{-1}). When photosystem II was set as the secondary objective, the highest fluxes through the photosystem II reaction were given by the phosphate limitation, mixotrophic, and low salinity conditions (0.016 mmol gDW^{-1} h^{-1}).

Lack of light is likely to be the greatest contributing factor to decrease in growth as low oxygen concentration does not seem to stunt growth, as the proportional decrease in biomass was lower relative to the standard control conditions. On the other hand, there appears to be little to no flux for the biomass or the photosystem I reaction in the dark conditions. This is supported by Vu et al. (2013), who reported that lower yields under dark conditions may be due to the limited generation of energy (ATP) and reductant (NADPH) from glycogen in the absence of photoautotrophic growth. When optical density was varied through the batch growth conditions (OD 0.4, 1.0, 3.0, and 5.0), the transmission of light through the cultures decreased as the dry cell weight (DCW) increased. Equal reduction in transcript levels for the photosynthetic apparatus was previously observed in all macronutrient-limited conditions studied (Ludwig and Bryant, 2012a). We infer that heat shock, mixotrophic growth, and phosphate limitation have the largest effect on reducing growth rate, as there was a complete impairment of biomass production predicted across all of our objectives for these conditions. This is in line with reported findings (Ludwig and Bryant, 2012a), where perturbations caused by phosphate limitation had a greater impact on the global transcription pattern than observed for high irradiance or dark treatments.

Synechococcus sp. PCC 7002 is known to possess one of the greatest tolerances for high light intensity among cyanobacteria (with an upper limit of approximately 2,000 μmol photons m^{-2} s^{-1}) (Xiong et al., 2015). This was evident from our predictions for all three pairs of objectives, where flux through the biomass pathway during high light intensity was slightly higher than the control condition (0.192 mmol gDW^{-1} h^{-1} compared with 0.053 mmol gDW^{-1} h^{-1}). Although the fluxes through photosystem II were disrupted, the fluxes through photosystem I were still maintained (0.058 mmol gDW^{-1} h^{-1} in the high light intensity condition as opposed to 0.016 mmol gDW^{-1} h^{-1} in the control). Heat shock resulted in no fluxes through any of the four reactions within all three objective pairs. It was previously reported that transcript levels for genes encoding photosystem I decreased slightly in cells grown at high salinity and remained constant at low salinity (Ludwig and Bryant, 2012b). On the other hand, it was found that transcript levels for genes encoding photosystem II did not change in response to fluctuations in salinity (Ludwig and Bryant, 2012b).

Multi-omic Principal-Component Analysis

It can be argued that analyzing single-omic data alone has limited relevance in the context of metabolic processes, as it does not capture the full complexity of the phenotype in relation to environmental variability. The hybrid approach proposed in this work connects transcriptomic and fluxomic data using a data-driven multi-view approach that supports machine learning algorithms to yield more accurate predictions (Culley et al., 2020). Considering the vast dimensionality of multi-omic models, the identification of biologically meaningful information can prove to be challenging. As a non-parametric statistical technique, principal-component analysis (PCA) was incorporated into our workflow for identifying patterns and genes/reactions responsible for the most variance in the datasets (Brunk et al., 2016).

The PCA indicates the proportion of variance exhibited by fluxes in the first two dimensions for each objective pair. For all three pairs of objectives in Figures 2A–2C, over 68.31% of the variance can be explained by

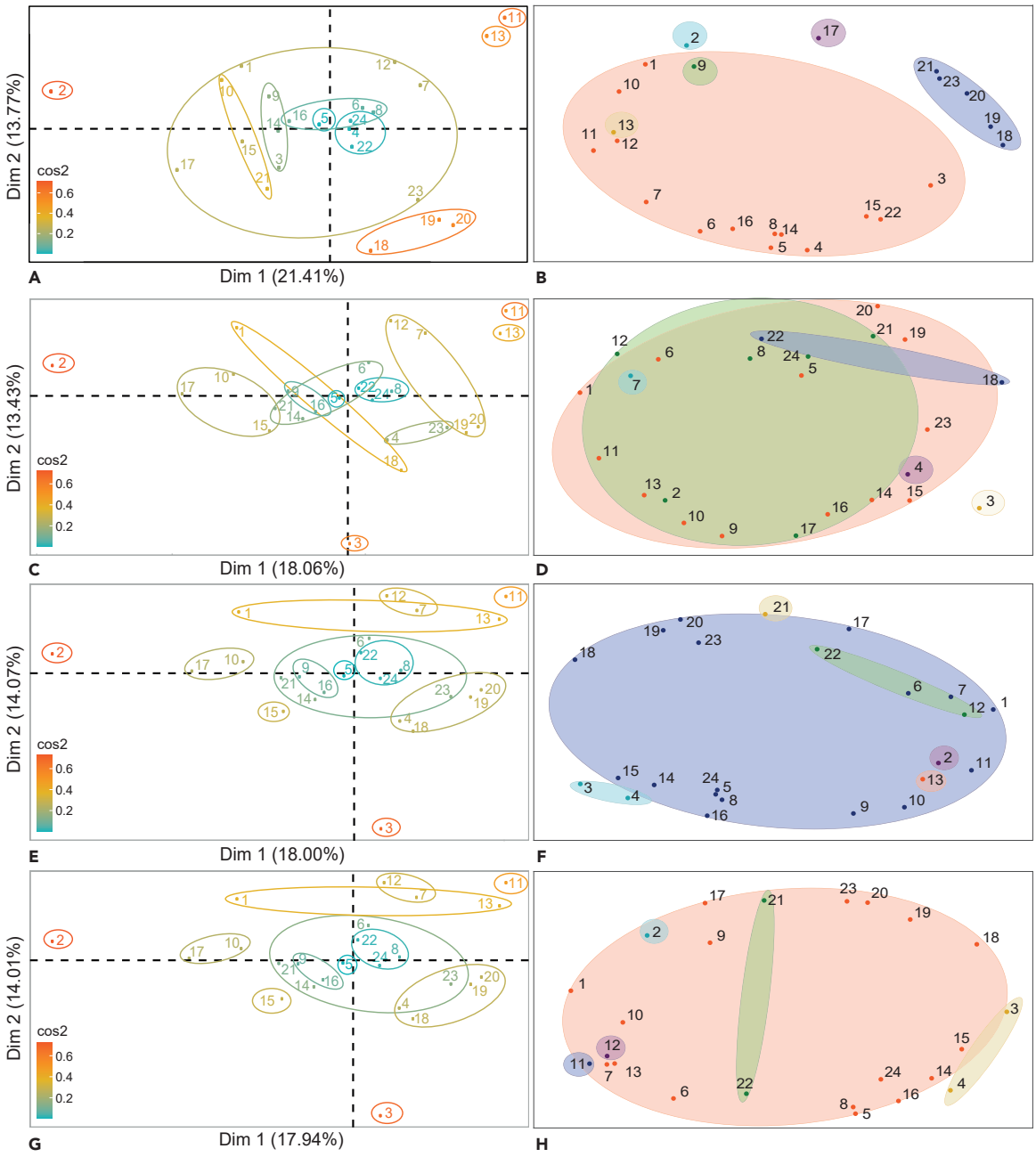


Figure 3. Transcriptomic and Multi-omic Principal-Component Analysis (PCA) and *k*-means

PCA and *k*-means clustering conducted for 3,187 gene transcripts (A and B) and both transcripts and fluxes, with Biomass-ATP maintenance (C and D), Biomass-Photosystem I (E and F), and Biomass-Photosystem II (G and H) as objective pairs. In the PCA plots (A, C, E, and G), growth conditions are colored according to their cos2 value, indicating the contribution of the first two components to the squared distance of each condition to the origin (Abdi and Williams, 2010). For *k*-means, data are clustered by condition (where the color of the ellipses represents different clusters) and the number of clusters was determined following silhouette analysis (k=6). When compared to using the transcriptomic dataset alone, the combined proportion of variance for PCA in the first two dimensions was slightly higher when gene transcript data was used in isolation than when it was combined with fluxes. For *k*-means clustering, the change in objective pair used for FBA did not result in a significant difference in the clusters formed. However, there is a clear demarcation between clusters of conditions that limit growth (e.g., low light, sulfate limitation) and those that promote growth (e.g., high light, nitrate supplementation).

the first dimension when considering flux data alone. As shown in plot (a), the high light intensity condition contributed the highest score for the first dimension, accounting for the vast majority of the variance. On the other hand, plots (b) and (c) showed that high light intensity, phosphate limitation, mixotrophic, and low salinity were the highest scoring conditions in the first dimension. For the second dimension, the highest score was given by high salinity, iron limitation, urea, 30°C and oxidative stress in plot (a), and high light intensity, phosphate limitation, mixotrophic, and low salinity conditions in plots (b) and (c).

When considering only the gene transcript data (Figure 3A), the combined proportion of variance that could be accounted for by the first two dimensions was vastly reduced (only 35.18% compared with 75.70%–86.27% for flux data). The conditions with the largest scores for the first dimension were sulfate and iron limitation, followed by oxidative stress, 30°C, and high salinity. Once again, sulfate and iron limitation were the highest in the second dimension, along with phosphate limitation, nitrogen limitation, dark anoxic/oxic, and the last phase of the batch growth (OD 5.0).

When using a combined dataset of both gene transcripts and fluxes (Figures 3C, 3E, and 3G), the total proportion of variance that could be explained in two dimensions for all three objective pairs was lower than using transcript data alone (31.43%–32.07%). The highest scores were given by iron and sulfate limitation in the first dimension and by dark oxic, dark anoxic, OD 5.0, iron limitation, and sulfate limitation in the second dimension.

A full list of gene transcript and calculated fluxes are included in Data S1. For a list of the top 10 contributions of genes and reactions to the principal components, we refer the reader to the Supplemental Information (Tables S3–S5).

Pathway-Level Analysis of Principal Components

To further examine the most metabolically significant pathways or cellular processes, we also performed a pathway-level PCA while categorizing genes and reaction by their main function. Owing to the varying number of reactions within each pathway, both the pathway sum and average contribution to the variance from the first two principal components were calculated. In Figures 4A and 4B, the sum of all contributions to variance within each pathway or COG (Cluster of Orthologous Groups) category is summarized. For the gene transcripts (Figure 4A), the COGs with the highest sum of variance within the first two principal components were poorly characterized (with general or unknown function). It can be observed that for each pair of flux objectives in (Figure 4B), the pathways that contribute the most to the first and second components were similar: cofactor and vitamin metabolism, nucleotide metabolism, energy metabolism, lipid metabolism, amino acid metabolism, carbohydrate metabolism, and transport metabolism. These pathways can be directly linked to cellular growth because many of their products are biomass precursors or compounds that can be catabolized to produce energy, i.e., carbohydrates, proteins, and fats.

The radar plots in Figures 4C–4F depict the average contributions to the variance within each pathway for the first and second principal components. The average contribution was higher for reactions (fluxes) than genes because the number of genes in each COG category was greater than the number of reactions in each subsystem of the GSMM. The pathways with the highest average contributions for gene transcripts were nucleotide/amino acid metabolism in the first component and chromatin structure and dynamics in the second component. For all three objective pairs, amino acid, aminoacyl-tRNA, and peptidoglycan biosyntheses were relevant in the first component (with an average contribution between 0.26 and 0.3). For the second component, the pathways with the largest contribution varied for each objective. Coenzyme and thiamine metabolism both had an average contribution greater than 0.45 for the second component in relation to ATP and photosystem I fluxes, with the addition of hydrogen metabolism for ATP and pyridine metabolism for photosystem I. On the other hand, purine and nucleotide metabolism had the highest contributions in the second component for photosystem II. However, many of these pathways contained only one or two reactions, which caused these results to be skewed; for example, the purine metabolism pathway had an average contribution of 1.36, but only contains one nucleotide phosphodiesterase reaction (PDE2).

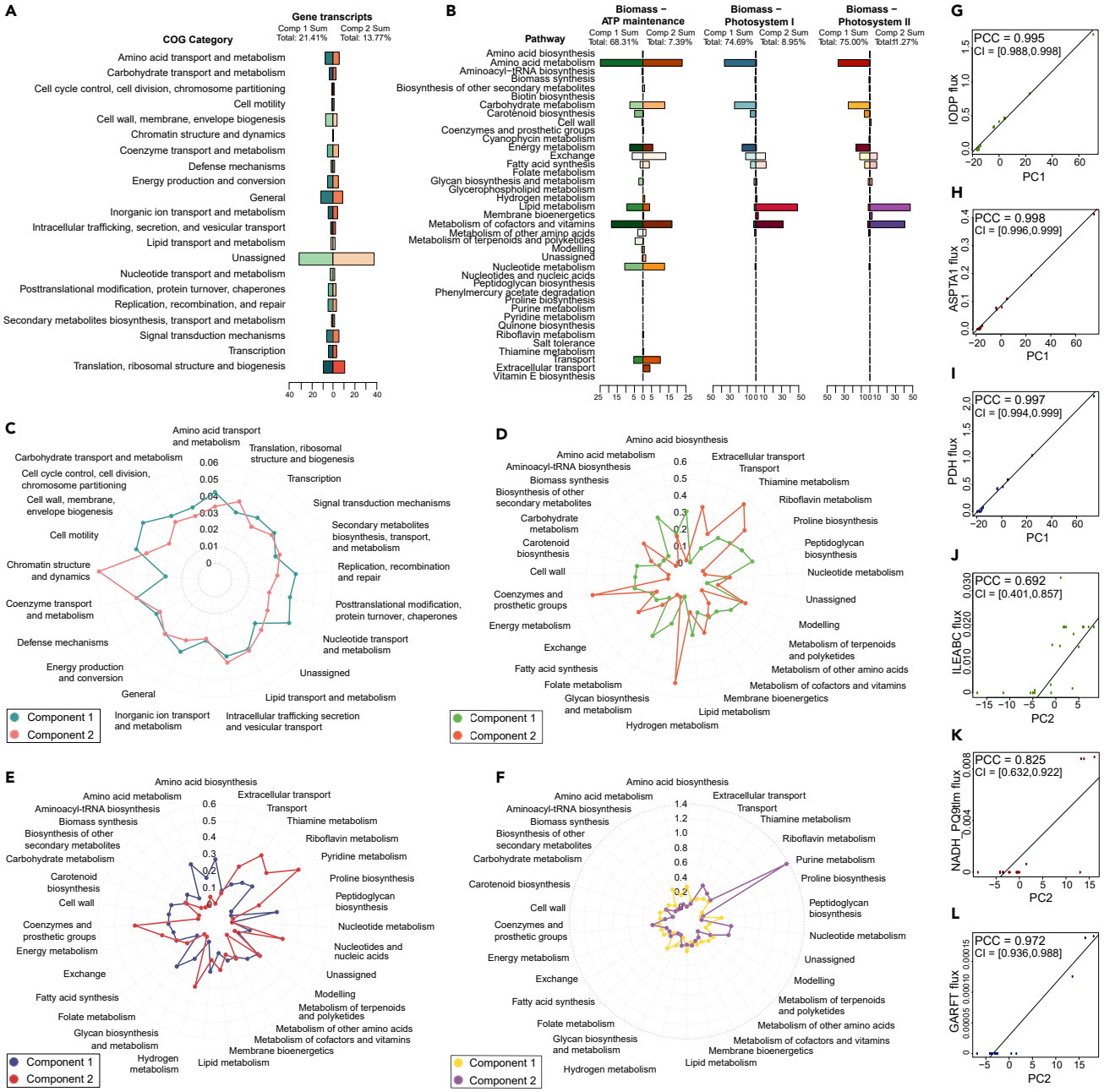


Figure 4. Pathway-Based PCA to Identify Pathway and Reaction Contribution to Variance across Conditions

(A and B) Component sums by pathway. Principal component contributions summed across reactions within each COG category/pathway, to decompose the metabolic function of the main contributors to variance. The total percentage of contribution to variance of the first two principal components is also given for each dataset: (A) gene transcripts and (B) fluxes calculated with objective functions Biomass-ATP maintenance, Biomass-Photosystem I, and Biomass-Photosystem II.

(C–F) Component contribution by pathway. Average principal component contributions within each pathway, calculated across all gene transcripts (C) and fluxes for each objective function pair: (D) Biomass-ATP maintenance, (E) Biomass-Photosystem I, and (F) Biomass-Photosystem II.

(G–L) Interpreting PCA coordinates with fluxes. The Pearson correlation coefficient (PCC) was calculated between the principal component coordinates and the flux values across the conditions in three pairs of objectives. The coordinates for the first principal component (*x* axis) and flux (*y* axis) across the 24 conditions were plotted for the following reactions: (G) inorganic diphosphatase (IODP) for Biomass-ATP maintenance, (H) aspartate transaminase (ASPTA1) for Biomass-Photosystem I, and (I) pyruvate dehydrogenase (PDH) for Biomass-Photosystem II. For the second principal component, the reactions were: (J) L-isoleucine transport via ABC system (ILEABC) for Biomass-ATP maintenance, (K) NADH dehydrogenase type II in the thylakoid membrane (NADH_PQ9tlm) for Biomass-Photosystem I, and (L) phosphoribosylglycinamide formyltransferase (GARFT) for Biomass-Photosystem II. The PCC with their respective 95% confidence intervals (CI) are displayed within each plot.

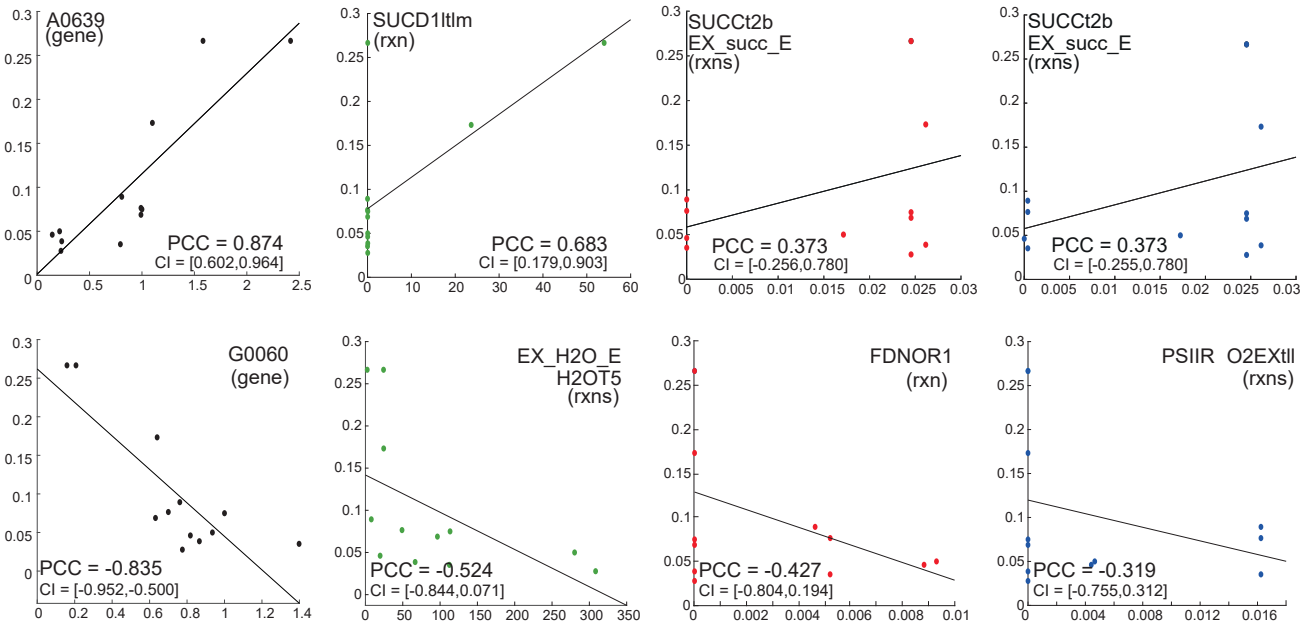


Figure 5. Top PCCs with Their Respective 95% Confidence Intervals (CI) between Gene Transcript/Reaction Flux Data (*x* axis) and Growth Rates (*y* axis)

Left to right: Black, top correlated genes; green, top correlated reactions when maximizing Biomass-ATP maintenance flux; red, top correlated reactions when maximizing Biomass-Photosystem I flux; blue, top correlated reactions when maximizing Biomass-Photosystem II flux. Tables S9–S13 list the top 10 genes/reactions in the dataset that are positively or negatively correlated with growth and their respective PCC values. Additional figures of top 10 genes/reactions and their respective PCC values are provided in the Supplemental Information (Figures S1–S4).

Finally, to characterize the PCA in the context of single reactions, we analyzed the principal component coordinates for all growth conditions against different reaction fluxes selected from the top 10 contributors to the variance in each of the three objective pairs (see Figures 4G–4L). These plots confirmed that a large part of the variance could be explained in the first principal component, as the first component coordinates showed a near-perfect correlation with flux (> 0.99). The second principal component displayed a less consistent but still strongly positive correlation between coordinates and flux values. The set of reactions with the highest contributions to variance in the first and second components were completely different, but three main functional categories could be identified among these reactions. IODP and GARFT can be linked to nucleotide metabolism, ASPTA1 and ILEABC to amino acid metabolism, and PDH and NADH_PQ9tlm to energy metabolism.

Clustering

k-means is a clustering algorithm that computes clusters while iteratively minimizing the sum of squared Euclidean distances between each observation and its respective cluster mean (McLachlan et al., 2008). To assess whether the generated multi-omic datasets could identify clusters of growth conditions according to the respective omic responses, we applied *k*-means to the set of 24 growth conditions, considering gene expression, flux rates, and the combined expression/flux dataset.

For the flux data (plots d–f in Figure 2), the partitioning of *k*-means clusters varied depending on the pair of FBA objectives for which fluxes were calculated. Following silhouette analysis, the number of clusters set for plots (d–f) in Figures 2, 3B, 3D, 3F, and 3H was k=6. The full list of members of each cluster is reported in the Supplemental Information.

When combining both transcript and flux data (Figure 3D, 3F, and 3H), the clusters formed were less distinct. This suggests that fluxes could help to contribute more biological insights into metabolic reactions (through the metabolic network) that are not available in the transcriptomic data. Nevertheless, through the *k*-means analysis with transcripts-only and the combined multi-omic dataset of transcripts and fluxes,

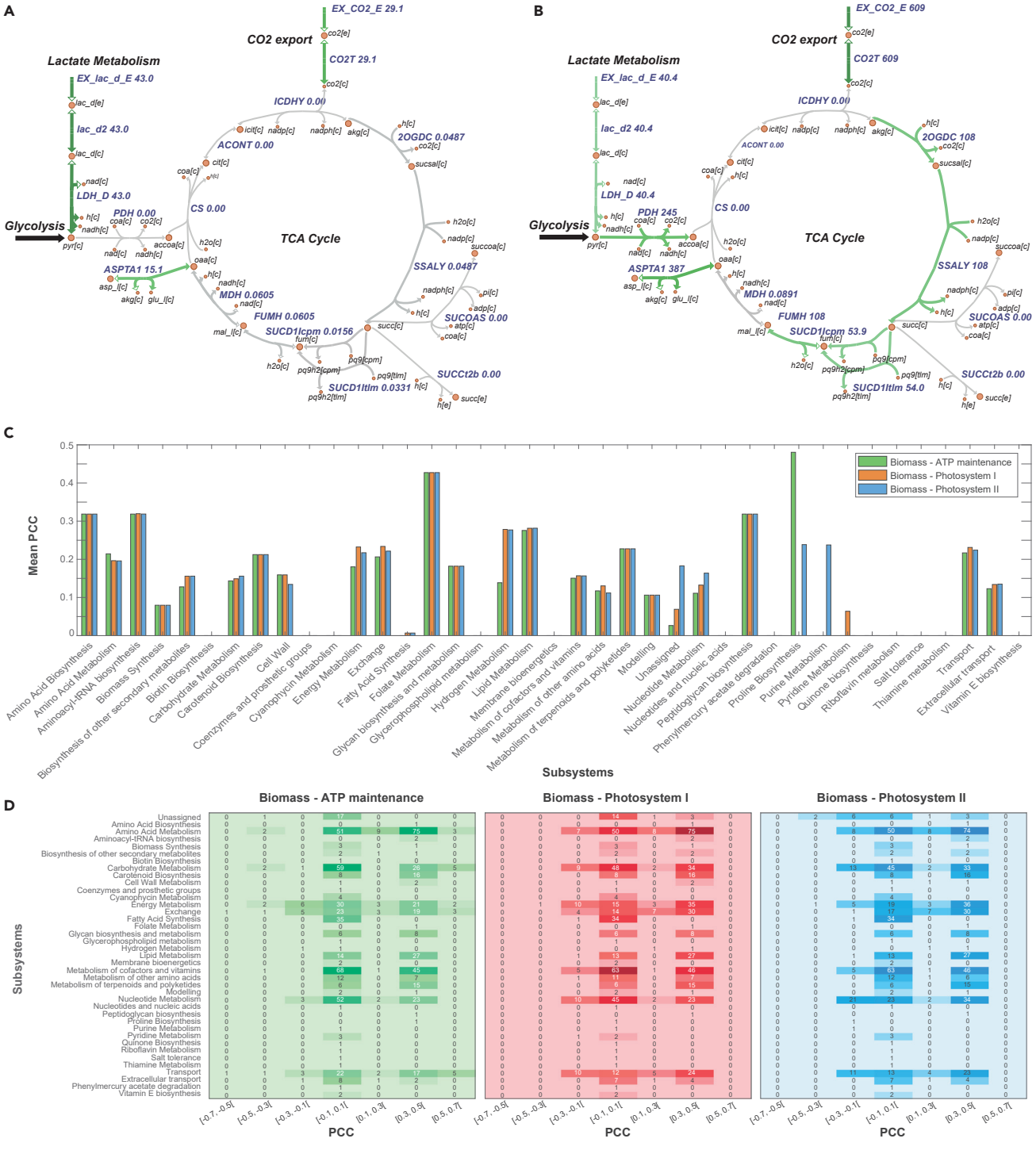


Figure 6. Flux Map Comparison and PCC Values by Subsystem

Comparison between flux values in the TCA cycle for (A) nitrogen limitation and (B) urea supplementation. The SUCD1ltlm and SUCD1lcpm reactions encoding succinate dehydrogenase were identified as having a strong positive correlation with the growth rate for the Biomass-ATP maintenance objective pair.

(C) Mean absolute Pearson correlation coefficient (PCC) values calculated between 12 experimental growth rates and their corresponding condition-specific GSMM reaction fluxes within each metabolic subsystem/pathway in the *Synechococcus* sp. PCC 7002 GSMM. The highest mean absolute correlations were identified for folate metabolism, proline, and amino acid biosynthesis.

Figure 6. *Continued*
(D) Reactions within each model subsystem sorted into classes of PCC values obtained between growth rates and flux rates in each objective pair (Biomass-ATP maintenance, Biomass-Photosystem I, Biomass-Photosystem II). The Biomass-ATP maintenance pair yielded the highest positive PCC values [0.5, 0.7[for reactions within the carbohydrate, amino acid, energy, transport, and exchange metabolic pathways.

we conclude that clustering techniques benefit from analyzing the flux and transcript datasets in isolation rather than combining them, as this avoids an increase in data dimensionality that cannot be easily reduced.

In most of the *k*-means plots in Figure 3, changing the objective pair used for FBA did not result in a significant difference in the clusters formed. However, there was sometimes a demarcation between clusters of conditions that limited growth (e.g., low light, nutrient limitation) and those that promoted growth (e.g., high light availability, nutrient supplementation). In some instances, certain conditions (e.g., heat shock, high light intensity, iron limitation) were isolated within a single cluster. Furthermore, reducing the number of dimensions in the data following PCA could serve to reduce noise and make the definition of clusters even clearer.

LASSO Regression

Regression-based algorithms can ascertain a mapping function, given a number of continuous output variables *y* and a number of real-valued or discrete input variables *x*. For each input variable, a coefficient is estimated with linear regression, determining the importance of the input variable toward predicting the output variable. In our study, the least absolute shrinkage and selection operator (LASSO) was used to select a subset of the input variables by minimizing the number of nonzero coefficients. It employs L1 regularization, which penalizes the sum of absolute values of all the coefficients; this sets the coefficients of unnecessary or recursive features equal to zero, resulting in a sparser matrix (Tibshirani, 1996). The regularization enables the identification of important predictors, elimination of redundant predictions, and generation of shrinkage estimates with lower predictive errors.

The LASSO regression identified genes and reactions in the model that are strongly related to *in vivo* growth rates through the retention of non-zero predictor coefficients (see Table S2). The full calculation of LASSO coefficients is provided in Data S2.

A complete list of non-zero predictors retained by LASSO for each dataset is given in the Supplemental Information. The functional classifications were provided by CyanOmics in the case of genes (CY Category and CY Sub Category) or the *subsystems* field within the model GSMM for reactions. In Tables S6 and S8 (where both transcript and flux datasets or only the gene transcripts are considered), the same non-zero predictors (genes) were retained, irrespective of the objective pair used for FBA. The genes yielding positive coefficients were associated with photosynthesis and respiration or post-translational modification of proteins. When applying the LASSO algorithm to the flux-only dataset (see Table S7), the non-zero coefficients retained were primarily related to the metabolism of nucleotides, cofactors, and vitamins and pathways relating to energy generation, such as carbohydrate and amino acid metabolic pathways. Such co-factors are often composed of metal ions, for which there are numerous transport and exchange reactions, e.g., cobalt and manganese.

Correlation Analysis

The Pearson correlation coefficients were calculated to ascertain the strength of the association between transcripts and/or flux rates and growth across different conditions (see Table S2). The absolute Pearson correlation coefficients were sorted in descending order, and the top 10 positive/negative correlation coefficients for each dataset are listed in the Supplemental Information. Figure 5 shows the highest positive and negative Pearson correlation coefficients for transcript- and flux-only datasets. The gene A0639 encodes a phycocyanin-associated phycobilisome rod-core linker polypeptide, which is an important component of the photosynthetic apparatus. This confirms that photosynthesis and energy metabolism are directly correlated with cellular growth.

For the Biomass-ATP maintenance flux objectives, the selection of reactions encoding succinate dehydrogenase in the cytoplasmic and thylakoid membranes (SUCD1lcpm and SUCD1ltlm) demonstrated the importance of the tricarboxylic acid (TCA) cycle in the generation of energy for biomass accumulation.

To illustrate this, a comparison of flux values between the nitrogen-limited and urea-supplemented growth conditions is provided in Figures 6A and 6B. It can be seen in Table S2 that the limitation or supplementation of a nitrogen source had a direct effect on the growth rate. Cyanobacteria have long been known to possess a unique TCA cycle where an alternative reaction homologous to 2OGDH is used to convert alpha-ketoglutarate into succinyl semialdehyde (2OGDC), which is subsequently converted into succinate (via succinate-semialdehyde dehydrogenase, i.e., SSALY) (Zhang and Bryant, 2011; Steinhauser et al., 2012). 2OGDC and SSALY were found to carry negligible flux under phototrophic conditions in (Hendry et al., 2016), which was supported by the flux values derived for the standard control for our simulations (3.599 mmol gDW^{-1} h^{-1} for both reactions), as well as the growth-limiting conditions such as nitrogen limitation (Figure 6A). This suggests that succinate dehydrogenation plays an important role in growth; it is known to still take place in dark, anoxic conditions (McNeely et al., 2010) where there is an increased flux toward succinate during the dark period, driving ATP production through respiratory electron transport (Sarkar et al., 2019).

Figure 6C shows the mean absolute PCC values among reactions within each subsystem, whereas Figure 6D shows the number of reactions within a given range of PCC values for each subsystem/pathway listed in the model, to account for the differing number of reactions in each pathway in the model. The pathway with the largest mean absolute correlation across all the flux objectives was the folate metabolism. In *Synechocystis* sp. PCC 6803, folate is synthesized from chorismate and is known to be important for cellular processes such as DNA replication, repair, and methylation, in addition to being a vital precursor for the biosynthesis of certain amino acids, co-factors, nucleotides, and tRNAs (Mills et al., 2020). The highest mean PCC in the biomass-ATP maintenance flux pair corresponded with the proline biosynthesis pathway/reaction. Interestingly, it has been found that proline accumulation is highly induced in stress conditions in cyanobacteria, especially high salinity (increased NaCl concentration) because it plays a role in osmoprotection, antioxidative defense, and signaling (Hayat et al., 2012; Pingkhanont et al., 2019). Across all flux objectives, the majority of reactions in Figure 6D have a correlation value between [-0.1, 0.1[or [0.3, 0.5[, implying most reactions had little to no significant correlation or a moderately positive correlation. The strongest positive correlation values [0.5, 0.7[were found between the Biomass-ATP maintenance flux pair and the growth rates. It can be seen that these reactions were classified under various pathways, i.e., carbohydrate, amino acid, energy, and transport- and exchange-related metabolism.

DISCUSSION

In this work, we showed how using a hybrid multi-view approach with multi-omic data and machine learning to yield metabolically significant fluxes enabled the identification of trends in data that were not apparent using solely transcriptomic data. We used condition-specific FBA to obtain flux distributions with L2-regularized bilevel optimization.

The flux distributions obtained for four key reactions showed clear differences in pathway activity across the various conditions and also between the three pairs of objectives used. When comparing the results across the types of datasets used, it is clear that complex metabolic and phenotypic outcomes as a result of adaptation to a changing environment are difficult to predict from gene expression alone. Condition-specific metabolic models within a machine learning framework allowed for the detection of coordinated responses shared between different data types, as well as the variation in responses across different growth conditions. Although a large number of studies express the maximization of biomass as the only objective when performing FBA, it is imperative to recognize that in reality most organisms have multiple objectives to satisfy. It has been well established that the activity of biosynthetic and energy-generating pathways increases with the growth rate (Bernstein et al., 2014), which led us to implement multi-level regularized optimization in our pipeline, considering more than one objective function.

Specifically, when calculating the flux distribution across conditions, biomass was chosen as the primary objective, whereas the secondary objective was set to ATP maintenance, photosystem I, or photosystem II, to reflect the main cellular goals of cyanobacteria. Biomass was chosen as a primary objective to represent the maximization of growth rate and cellular yields (Feist and Palsson, 2010; Yuan et al., 2016; Lakshmanan et al., 2019), which is a critical consideration for the production of biofuels by cyanobacteria as this informs the substrate uptake rates and maintenance requirements that indicate fundamental cellular growth requirements. The chosen secondary objectives are key pathways involved in energy metabolism

during photosynthesis. Simulating the cost of ATP maintenance can help to examine the energy required for sustaining metabolic activity even in the absence of growth. The incorporation of the photoexcitation reactions occurring within photosystems I and II served to characterize how flux under various conditions reflects the light harvesting and energy transfer via photon absorption through these complexes. Thus, solving the quadratic optimization problem for multiple pairs of objectives helped to resolve trade-offs by considering the conditions and constraints affecting each of these objectives (Sajitz-Hermstein and Nikoloski, 2016; Occhipinti et al., 2020).

Our results suggest that it is worth using model-generated flux data that incorporates transcriptomics to conduct machine learning analyses. The flux data were initially informed by transcriptomic data as the condition-specific gene expression profiles were generated by combining them with a baseline GSMM for *Synechococcus* sp. PCC 7002; in this way, gene transcripts already constituted an important component of the FBA. Furthermore, reducing the number of dimensions in the data following PCA can serve to reduce noise and make the definition of clusters even clearer. In addition to this, a reduced set of predictors were identified as being related to growth as a result of the LASSO regularization. Specifically, the identification of reactions by LASSO as key features, which are of potential use for the prediction of growth rates, supports the inclusion of metabolic fluxes as features for future applications of regression techniques with smaller, more concise sets of flux data.

The reactions identified as being strongly correlated with the growth rate in the flux datasets (SUCD1Itlm, SUCD1Icpm, ME2) suggest that fluxes can help to gain more biological insights into machine learning analyses. As a different, unrelated set of genes displayed a strong correlation with the growth rate, it is evident that analyzing both transcriptomic and fluxomic data provides a more complete picture of cyanobacterial metabolism than single-omic analyses. In particular, the role of metal transport pathways in cyanobacteria was significant because they are highly relevant in the context of photosynthesis. The detection of latent, biologically significant patterns and adaptive mechanisms to fluctuations in light intensity and salinity elucidates the maintenance of metabolic efficiency at the cellular level, as well as the attainment of multiple cellular objectives.

Algal engineering supplemented with data from multi-omic studies can contribute to informing the scale-up of these processes. Such multi-omic data are sensitive enough to detect the effect of stress on metabolism. Metabolic engineers could apply this pipeline to test more strategies *in silico* when developing the optimal production host, or to analyze multi-omic outputs (both independently and in combination with other omic data). In this regard, the use of transcriptomic data to characterize fluxomic predictions elucidates many of the unique mechanisms employed by *Synechococcus* sp. PCC 7002 when adapting to changes in light intensity, salinity, and other conditions. In the case of cyanobacteria, we also emphasize the importance of assessing model inputs in accordance with specific growth conditions before conducting FBA. These contribute to the organism's underlying objective of maintaining metabolic efficiency for phototrophic growth and light-dependent photosynthesis. As a result of predicting and classifying metabolic profiles in various growth conditions, our approach sheds light on the cross-omic mechanisms of its adaptation process, which enables survival across a wide range of environments and stress conditions.

Limitations of the Study

The availability of exact measurements for the various growth conditions could yield more precise flux predictions. For example, in our case, the exact photon absorbance of the *Synechococcus* sp. PCC 7002 cultures was not available. Hence, the photon uptake constraints were approximated using DCW and photon consumption based on the availability of light. Likewise, the setting of nutrient uptake rates was approximated based on data provided by *in vivo* experiments rather than measured directly.

Furthermore, in this study we adopted linear transformations and linear methods, where possible. This was with the goal of maximizing the biological interpretability of the predictions, using quadratic terms for regularization only. However, different dimensionality reduction or clustering methods could be implemented, e.g., to elucidate any further non-linear relationship among the omic elements.

Finally, there is further potential for other types of omic data to be integrated into the model (e.g., from proteomic or metabolomic datasets). It is expected that integrating further omic datasets, i.e., further data views in our multi-view machine learning setting, could produce even more detailed insights into metabolic adaptations, or better support existing findings derived from transcriptomic and fluxomic data.

Resource Availability

Lead Contact

Dr. Claudio Angione, email: c.angione@tees.ac.uk.

Materials Availability

The study did not generate new unique reagents or other materials.

Data and Code Availability

The complete source code of our pipeline is freely available on GitHub at https://github.com/Angione-Lab/Synechococcus7002-metabolic-modelling.

The RNA-seq data integrated into the model were from transcriptomic datasets uploaded by Ludwig and Bryant (2011, 2012a, 2012b). These were acquired directly from the Cyanomics dataset (Yang et al., 2015), but have since been made available on the NCBI Sequence Read Archive. The accession numbers for the RNA-seq data reported in this paper are SRA:SRP007372, SRA:SRP013965 and SRA:SRP066851.

METHODS

All methods used in this study can be found in the accompanying Transparent Methods section in the Supplemental Information file.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.isci.2020.101818.

ACKNOWLEDGMENTS

We would like to thank Dr. John Hendry from Penn State University for providing advice on fixing boundary constraints for the *Synechococcus* sp. PCC 7002 GSMM. We would also like to thank Dr. Guido Zampieri from the University of Padova for discussions on the methodology. C.A. and P.K.S.M.R. received funding from BBSRC grant CBMNet-PoC-D0156. P.K.S.M.R. would like to thank Research England for funding support through the Expanding Excellence in England (E3) scheme. C.A. would like to acknowledge the support from UKRI Research England's THYME project.

AUTHOR CONTRIBUTIONS

Conceptualization, S.V. and C.A.; Data Curation, S.V.; Formal Analysis, S.V.; Funding Acquisition, P.K.S.M.R. and C.A.; Investigation, S.V.; Methodology, S.V and C.A.; Project Administration, C.A.; Resources, P. K.S.M.R. and C.A; Software, S.V. and C.A.; Supervision, C.A.; Validation, S.V.; Visualization, S.V. and C.A.; Writing – Original Draft, S.V., P. K.S.M.R., and C.A.; Writing – Reviewing and Editing, S.V. and C.A.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: May 26, 2020
Revised: October 23, 2020
Accepted: November 13, 2020
Published: December 18, 2020

REFERENCES

Abdi, H., and Williams, L.J. (2010). Principal component analysis. Wiley Interdiscip. Rev. Comput. Stat. *2*, 433–459.

Abedpour, N., and Kollmann, M. (2015). Resource constrained flux balance analysis predicts selective pressure on the global structure of metabolic networks. BMC Syst. Biol. *9*, 88.

Ahmad, A., Pathania, R., and Srivastava, S. (2020). Biochemical characteristics and a genome-scale metabolic model of an indian euryhaline cyanobacterium with high polyglucan content. Metabolites *10*, 177.

Angermayr, S.A., Rovira, A.G., and Hellingwerf, K.J. (2015). Metabolic engineering of cyanobacteria for the synthesis of commodity products. Trends Biotechnol. *33*, 352–361.

Angione, C. (2018). Integrating splice-isoform expression into genome-scale models characterizes breast cancer metabolism. Bioinformatics *34*, 494–501.

Angione, C. (2019). Human systems biology and metabolic modelling: a review—from disease metabolism to precision medicine. Biomed. Res. Int. *2019*, 8304260.

Angione, C., Costanza, J., Carapezza, G., Lió, P., and Nicosia, G. (2015). Multi-target analysis and design of mitochondrial metabolism. PLoS One *10*, e0133825.

van der Ark, K.C., van Heck, R.G., Dos Santos, V.A.M., Belzer, C., and de Vos, W.M. (2017). More than just a gut feeling: constraint-based genome-scale metabolic models for predicting functions of human intestinal microbes. Microbiome *5*, 78.

Babele, P.K., and Young, J.D. (2020). Applications of stable isotope-based metabolomics and fluxomics toward synthetic biology of cyanobacteria. Wiley Interdiscip. Rev. Syst. Biol. Med. *12*, e1472.

Bernstein, H.C., Konopka, A., Melnicki, M.R., Hill, E.A., Kucek, L.A., Zhang, S., Shen, G., Bryant, D.A., and Beliaev, A.S. (2014). Effect of mono-and dichromatic light quality on growth rates and photosynthetic performance of *Synechococcus* sp. pcc 7002. Front. Microbiol. *5*, 488.

Blanco-Ameijeiras, S., Moisset, S.A., Trimborn, S., Campbell, D.A., Heiden, J.P., and Hassler, C.S. (2018). Elemental stoichiometry and photophysiology regulation of *Synechococcus* sp. pcc 7002 under increasing severity of chronic iron limitation. Plant Cell Physiol. *59*, 1803–1816.

Brunk, E., George, K.W., Alonso-Gutierrez, J., Thompson, M., Baidoo, E., Wang, G., Petzold, C.J., McCloskey, D., Monk, J., Yang, L., et al. (2016). Characterizing strain variation in engineered e. coli using a multi-omics-based workflow. Cell Syst. *2*, 335–346.

Carroll, A.L., Case, A.E., Zhang, A., and Atsumi, S. (2018). Metabolic engineering tools in model cyanobacteria. Metab. Eng. *50*, 47–56.

Clark, R.L., McGinley, L.L., Purdy, H.M., Korosh, T.C., Reed, J.L., Root, T.W., and Pfleger, B.F. (2018). Light-optimized growth of cyanobacterial cultures: growth phases and productivity of biomass and secreted molecules in light-limited batch growth. Metab. Eng. *47*, 230–242.

Culley, C., Vijayakumar, S., Zampieri, G., and Angione, C. (2020). A mechanism-aware and multiomic machine-learning pipeline characterizes yeast cell growth. Proc. Natl. Acad. Sci. U S A *117*, 18869–18879.

Damini, J., Annesha, S., Shinjinee, S., Swati, M., Pakrasi, H.B., and Wangikar, P.P. (2020). A novel cyanobacterium *Synechococcus elongatus* pcc 11802 has distinct genomic and metabolomic characteristics compared to its neighbor pcc 11801. Sci. Rep. *10*, 191.

Dougherty, B.V., Moutinho, T.J., Jr., and Papin, J. (2017). Accelerating the Drug Development Pipeline with Genome-Scale Metabolic Network Reconstructions, *6* (Wiley-VCH Verlag GmbH & Co. KGaA).

Fatma, Z., Hartman, H., Poolman, M.G., Fell, D.A., Srivastava, S., Shakeel, T., and Yazdani, S.S. (2018). Model-assisted metabolic engineering of *Escherichia coli* for long chain alkane and alcohol production. Metab. Eng. *46*, 1–12.

Feist, A.M., and Palsson, B.O. (2010). The biomass objective function. Curr. Opin. Microbiol. *13*, 344–349.

Gunde-Cimerman, N., Plemenitaš, A., and Oren, A. (2018). Strategies of adaptation of microorganisms of the three domains of life to high salt concentrations. FEMS Microbiol. Rev. *42*, 353–375.

Haas, R., Zelezniak, A., Iacovacci, J., Kamrad, S., Townsend, S., and Ralser, M. (2017). Designing and interpreting 'multi-omic'experiments that may change our understanding of biology. Curr. Opin. Syst. Biol. *6*, 37–45.

Hayat, S., Hayat, Q., Alyemeni, M.N., Wani, A.S., Pichtel, J., and Ahmad, A. (2012). Role of proline under changing environments: a review. Plant Signal. Behav. *7*, 1456–1466.

Hendry, J.I., Bandyopadhyay, A., Srinivasan, S., Pakrasi, H.B., and Maranas, C.D. (2020). Metabolic model guided strain design of cyanobacteria. Curr. Opin. Biotechnol. *64*, 17–23.

Hendry, J.I., Gopalakrishnan, S., Ungerer, J., Pakrasi, H.B., Tang, Y.J., and Maranas, C.D. (2019). Genome-scale fluxome of *Synechococcus elongatus* utex 2973 using transient 13c-labeling data. Plant Physiol. *179*, 761–769.

Hendry, J.I., Prasannan, C., Ma, F., Möllers, K.B., Jaiswal, D., Digmurti, M., Allen, D.K., Frigaard, N.U., Dasgupta, S., and Wangikar, P.P. (2017). Rerouting of carbon flux in a glycogen mutant of cyanobacteria assessed via isotopically non-stationary 13c metabolic flux analysis. Biotechnol. Bioeng. *114*, 2298–2308.

Hendry, J.I., Prasannan, C.B., Joshi, A., Dasgupta, S., and Wangikar, P.P. (2016). Metabolic model of *Synechococcus* sp. pcc 7002: prediction of flux distribution and network modification for enhanced biofuel production. Bioresour. Technol. *213*, 190–197.

Hitchcock, A., Hunter, C.N., and Canniffe, D.P. (2020). Progress and challenges in engineering cyanobacteria as chassis for light-driven biotechnology. Microb. Biotechnol. *13*, 363–367.

Huang, Z., Lee, D.Y., and Yoon, S. (2017). Quantitative intracellular flux modeling and applications in biotherapeutic development and production using cho cell cultures. Biotechnol. Bioeng. *114*, 2717–2728.

Jagadevan, S., Banerjee, A., Banerjee, C., Guria, C., Tiwari, R., Baweja, M., and Shukla, P. (2018). Recent developments in synthetic biology and metabolic engineering in microalgae towards biofuel production. Biotechnol. Biofuels *11*, 185.

Jaiswal, D., Sengupta, A., Sohoni, S., Sengupta, S., Phadnavis, A.G., Pakrasi, H.B., and Wangikar, P.P. (2018). Genome features and biochemical characteristics of a robust, fast growing and naturally transformable cyanobacterium *Synechococcus elongatus* pcc 11801 isolated from India. Sci. Rep. *8*, 1–13.

Kashaf, S.S., Angione, C., and Lió, P. (2017). Making life difficult for clostridium difficile: augmenting the pathogen's metabolic model with transcriptomic and codon usage data for better therapeutic target characterization. BMC Syst. Biol. *11*, 25.

Lakshmanan, M., Long, S., Ang, K.S., Lewis, N.E., and Lee, D.Y. (2019). On the impact of biomass composition in constraint-based flux analysis. bioRxiv, 652040.

Luan, G., Zhang, S., and Lu, X. (2020). Engineering cyanobacteria chassis cells toward more efficient photosynthesis. Curr. Opin. Biotechnol. *62*, 1–6.

Ludwig, M., and Bryant, D.A. (2011). Transcription profiling of the model cyanobacterium *Synechococcus* sp. strain pcc 7002 by next-gen (solid™) sequencing of cdna. Front. Microbiol. *2*, 41.

Ludwig, M., and Bryant, D.A. (2012a). Acclimation of the global transcriptome of the cyanobacterium *Synechococcus* sp. strain pcc 7002 to nutrient limitations and different nitrogen sources. Front. Microbiol. *3*, 145.

Ludwig, M., and Bryant, D.A. (2012b). *Synechococcus* sp. strain pcc 7002 transcriptome: acclimation to temperature, salinity, oxidative stress, and mixotrophic growth conditions. Front. Microbiol. *3*, 354.

McLachlan, G.J., Bean, R.W., and Ng, S.K. (2008). Clustering. In Bioinformatics, M.K. Jonathan, ed. (Springer), pp. 423–439.

McNeely, K., Xu, Y., Bennette, N., Bryant, D.A., and Dismukes, G.C. (2010). Redirecting reductant flux into hydrogen production via metabolic engineering of fermentative carbon metabolism in a cyanobacterium. Appl. Environ. Microbiol. *76*, 5032–5038.

Mills, L.A., McCormick, A.J., and Lea-Smith, D.J. (2020). Current knowledge and recent advances in understanding metabolism of the model cyanobacterium *Synechocystis* sp. pcc 6803. Biosci. Rep. *40*, BSR20193325.

Montgomery, B.L. (2017). Seeing new light: recent insights into the occurrence and regulation of

chromatic acclimation in cyanobacteria. Curr. Opin. Plant Biol. *37*, 18–23.

Mukherjee, B., Madhu, S., and Wangikar, P.P. (2020). The role of systems biology in developing non-model cyanobacteria as hosts for chemical production. Curr. Opin. Biotechnol. *64*, 62–69.

Noreña-Caro, D., and Benton, M.G. (2018). Cyanobacteria as photoautotrophic biofactories of high-value chemicals. J. CO2 Util. *28*, 335–366.

O'Brien, E.J., Monk, J.M., and Palsson, B.O. (2015). Using genome-scale models to predict biological capabilities. Cell *161*, 971–987.

Occhipinti, A., Eyassu, F., Rahman, T.J., Rahman, P.K., and Angione, C. (2018). In silico engineering of *Pseudomonas* metabolism reveals new biomarkers for increased biosurfactant production. PeerJ *6*, e6046.

Occhipinti, A., Hamadi, Y., Kugler, H., Wintersteiger, C., Yordanov, B., and Angione, C. (2020). Discovering essential multiple gene effects through large scale optimization: an application to human cancer metabolism. IEEE/ACM Trans. Comput. Biol. Bioinform.

Oliver, N.J., Rabinovitch-Deere, C.A., Carroll, A.L., Nozzi, N.E., Case, A.E., and Atsumi, S. (2016). Cyanobacterial metabolic engineering for biofuel and chemical production. Curr. Opin. Chem. Biol. *35*, 43–50.

Opdam, S., Richelle, A., Kellman, B., Li, S., Zielinski, D.C., and Lewis, N.E. (2017). A systematic evaluation of methods for tailoring genome-scale metabolic models. Cell Syst. *4*, 318–329.

Pade, N., and Hagemann, M. (2014). Salt acclimation of cyanobacteria and their application in biotechnology. Life *5*, 25–49.

Pandhal, J., Noirel, J., Wright, P.C., and Biggs, C.A. (2009). A systems biology approach to investigate the response of synechocystis sp. pcc6803 to a high salt environment. Saline Syst. *5*, 1–16.

Pingkhanont, P., Tarasuntisuk, S., Hibino, T., Kageyama, H., and Waditee-Sirisattha, R. (2019). Expression of a stress-responsive gene cluster for mycosporine-2-glycine confers oxidative stress tolerance in *Synechococcus elongatus* pcc 7942. FEMS Microbiol. Lett. *366*, fnz115.

Randhawa, K.S., Relph, L.E., Armstrong, M.C., and Rahman, P.K. (2017). Biofuel production: tapping into microalgae despite challenges. Biofuels *8*, 261–271.

Rawat, I., Kumar, R.R., Mutanda, T., and Bux, F. (2013). Biodiesel from microalgae: a critical evaluation from laboratory to large scale production. Appl. Energ. *103*, 444–467.

Reed, J.L. (2012). Shrinking the metabolic solution space using experimental datasets. PLoS Comput. Biol. *8*, e1002662.

Reimers, A.M., Knoop, H., Bockmayr, A., and Steuer, R. (2016). Evaluating the stoichiometric and energetic constraints of cyanobacterial diurnal growth. arXiv. arXiv:1610.06859.

Ruffing, A.M., Jensen, T.J., and Strickland, L.M. (2016). Genetic tools for advancement of *Synechococcus* sp. pcc 7002 as a cyanobacterial chassis. Microb. Cell Fact. *15*, 190.

Rügen, M., Bockmayr, A., and Steuer, R. (2015). Elucidating temporal resource allocation and diurnal dynamics in phototrophic metabolism using conditional fba. Sci. Rep. *5*, 15247.

Sajitz-Hermstein, M., and Nikoloski, Z. (2016). Multi-objective shadow prices point at principles of metabolic regulation. Biosystems *146*, 91–101.

Sánchez, B.J., Zhang, C., Nilsson, A., Lahtvee, P.J., Kerkhoven, E.J., and Nielsen, J. (2017). Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints. Mol. Syst. Biol. *13*, 935.

Sarkar, D., Mueller, T.J., Liu, D., Pakrasi, H.B., and Maranas, C.D. (2019). A diurnal flux balance model of *Synechocystis* sp. pcc 6803 metabolism. PLOS Comput. Biol. *15*, e1006692.

Segre, D., Vitkup, D., and Church, G.M. (2002). Analysis of optimality in natural and perturbed metabolic networks. Proc. Natl. Acad. Sci. U S A *99*, 15112–15117.

Song, H.S., McClure, R.S., Bernstein, H.C., Overall, C.C., Hill, E.A., and Beliaev, A.S. (2015). Integrated in silico analyses of regulatory and metabolic networks of *Synechococcus* sp. pcc 7002 reveal relationships between gene centrality and essentiality. Life *5*, 1127–1140.

Song, H.S., Reifman, J., and Wallqvist, A. (2014). Prediction of metabolic flux distribution from gene expression data based on the flux minimization principle. PLoS One *9*, e112524.

Steinhauser, D., Fernie, A.R., and Araújo, W.L. (2012). Unusual cyanobacterial tca cycles: not broken just different. Trends Plant Sci. *17*, 503–509.

Tian, M., and Reed, J.L. (2018). Integrating proteomic or transcriptomic data into metabolic models using linear bound flux balance analysis. Bioinformatics *34*, 3882–3888.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. J. R. Stat. Soc. Ser. B Methodol. *58*, 267–288.

Toyoshima, M., Toya, Y., and Shimizu, H. (2020). Flux balance analysis of cyanobacteria reveals selective use of photosynthetic electron transport components under different spectral light conditions. Photosynth. Res. *143*, 31–43.

Vijayakumar, S., and Angione, C. (2017). Multi-omic data integration elucidates *Synechococcus* adaptation mechanisms to fluctuations in light intensity and salinity. In International Conference on Bioinformatics and Biomedical Engineering (Springer), pp. 220–229.

Vijayakumar, S., Conway, M., Lió, P., and Angione, C. (2017). Seeing the wood for the trees: a forest of methods for optimization and omic-network integration in metabolic modelling. Brief. Bioinformatics *19*, 1218–1235.

Vu, T.T., Hill, E.A., Kucek, L.A., Konopka, A.E., Beliaev, A.S., and Reed, J.L. (2013). Computational evaluation of *Synechococcus* sp. pcc 7002 metabolism for chemical production. Biotechnol. J. *8*, 619–630.

Wang, M., Luan, G., and Lu, X. (2019). Systematic identification of a neutral site on chromosome of *Synechococcus* sp. pcc 7002, a promising photosynthetic chassis strain. J. Biotechnol. *295*, 37–40.

Wang, Y., Chen, L., and Zhang, W. (2016). Proteomic and metabolomic analyses reveal metabolic responses to 3-hydroxypropionic acid synthesized internally in cyanobacterium synechocystis sp. pcc 6803. Biotechnol. Biofuels *9*, 209.

Włodarczyk, A., Selão, T.T., Norling, B., and Nixon, P.J. (2020). Newly discovered *Synechococcus* sp. PCC 11901 is a robust cyanobacterial strain for high biomass production. Communications Biology *3*, 1–14.

Wortel, M.T., Noor, E., Ferris, M., Bruggeman, F.J., and Liebermeister, W. (2018). Metabolic enzyme cost explains variable trade-offs between microbial growth rate and yield. PLoS Comput. Biol. *14*, e1006010.

Xiong, Q., Feng, J., Li, S.t., Zhang, G.y., Qiao, Z.x., Chen, Z., Wu, Y., Lin, Y., Li, T., Ge, F., et al. (2015). Integrated transcriptomic and proteomic analysis of the global response of *Synechococcus* to high light stress. Mol. Cell Proteomics *14*, 1038–1053.

Yang, Y., Feng, J., Li, T., Ge, F., and Zhao, J. (2015). Cyanomics: an integrated database of omics for the model cyanobacterium Synechococcus sp. pcc 7002. Database *2015*, bau127.

Yu, J., Liberton, M., Cliften, P.F., Head, R.D., Jacobs, J.M., Smith, R.D., Koppenaal, D.W., Brand, J.J., and Pakrasi, H.B. (2015). *Synechococcus elongatus* utex 2973, a fast growing cyanobacterial chassis for biosynthesis using light and co 2. Sci. Rep. *5*, 8132.

Yuan, H., Cheung, C., Hilbers, P.A., and van Riel, N.A. (2016). Flux balance analysis of plant metabolism: the effect of biomass composition and model structure on model predictions. Front. Plant Sci. *7*, 537.

Zampieri, G., Vijayakumar, S., Yaneske, E., and Angione, C. (2019). Machine and deep learning meet genome-scale metabolic modeling. PLoS Comput. Biol. *15*, e1007084.

Zhang, S., and Bryant, D.A. (2011). The tricarboxylic acid cycle in cyanobacteria. Science *334*, 1551–1553.

iScience, Volume *23*

Supplemental Information

A Hybrid Flux Balance Analysis and Machine Learning Pipeline Elucidates Metabolic Adaptation in Cyanobacteria

Supreeta Vijayakumar, Pattanathu K.S.M. Rahman, and Claudio Angione

1. Transparent Methods

1.1. Growth Conditions

A complete list of the culture conditions for *Synechococcus* sp. PCC 7002 is reported in Table S1, for which RNA sequencing data was integrated into GSMM to generate a condition-specific FBA framework.

1.2. Condition-Specific Metabolic Modeling

We initiated our pipeline by mapping the gene expression profiles for phototrophic growth in *Synechococcus* sp. PCC 7002 using multi-omic flux balance analysis and building condition-specific flux profiles, starting from a model recently published by Hendry et al. (2016).

Transcriptomic data were acquired in the form of RNA-Seq data from a series of studies previously conducted by Ludwig and Bryant (Ludwig and Bryant, 2011, 2012a,b). Such data were compiled in an online repository known as CyanOmics (Yang et al., 2015), an integrated omics analysis database containing omic data specific to *Synechococcus* sp. PCC 7002. Specific growth conditions are recorded in Table S1, along with a reference to the original paper for each condition. For each culture condition, we downloaded the reads assigned per kilobase of target per million mapped reads (RPKMs) as a measure of relative transcript abundance (Yang et al., 2015). Given this information, we calculated fold change values of each gene dividing the RPKM values under these conditions by the average expression of three standard control replicates for that gene.

In flux balance analysis (FBA), a steady-state is assumed to calculate all fluxes under time-invariance and spatial homogeneity for purposes of mass conservation (Heirendt et al., 2019). Mass-balance constraints are imposed on the GSMM to identify a range of points representing all feasible flux distributions. A phenotypic state in the solution space is then computed using linear programming with a set of values indicating the optimal conditions required to optimize a given cellular objective function (Ebrahim et al., 2016).

In our pipeline, we integrated condition-specific expression profiles generated from RNA sequencing data with the *Synechococcus* GSMM using METRADE (Angione and Lió, 2015). We then implemented a quadratic program for solving a regularized bi-level FBA, using Gurobi as a quadratic programming solver.

The regularized optimization problem was formulated as:

$$\begin{aligned} &\max && g^{\intercal}v - \frac{\sigma}{2}\, v^{\intercal}v \\ &\text{such that} && \max\, f^{\intercal}v, \quad Sv = 0, \\ &&& v^{\min}\varphi(\Theta) \leq v \leq v^{\max}\varphi(\Theta), \end{aligned} \tag{1}$$

where S is the stoichiometric matrix recording all reactions and metabolites in the *Synechococcus* GSMM, v is the vector of reaction flux rates. f and g are Boolean vectors of weights selecting the reactions in v in which flux rates are considered as the objectives: f selects the primary objective function (biomass) and g selects the secondary objective function (ATP maintenance, photosystem I or photosystem II). In order to obtain a unique flux solution, the outer level objective g is regularized by subtracting the concave function $\frac{\sigma}{2}$ $v^{\intercal}v$, where $\sigma = 10^{-6}$ (Heirendt et al., 2019). Upon solving this problem, a vector of fluxes (v) is obtained.

$v^{\min}$ and $v^{\max}$ are vectors representing the lower- and upper-limits for the flux rates in v for the unconstrained model. The gene set expression of reactions associated with the fluxes in v are represented by the vector Θ. φ is a function mapping the expression level of each gene set to a coefficient for the lower- and upper-limits of the corresponding reaction (Angione et al., 2016), and is defined as follows:

$$\varphi(\Theta) = [\mathbf{1} + \gamma\,|log(\Theta)|]^{\mathrm{sgn}(\Theta - \mathbf{1})}\,. \tag{2}$$

In this way, a specific RNA-Seq profile was mapped onto the GSMM for each growth condition, and solving the quadratic optimization problem yielded condition-specific flux rates. In the case of standard control flux, the expression vector was set to all ones; for all other conditions, RNA-Seq data was mapped to coefficients for the lower and upper limits of the corresponding reactions using the function in Equation 2. γ represents the strength of gene expression mapped to each reaction in the model (set to 3.5). A sensitivity analysis was conducted to assess the robustness of this parameter, whilst ensuring maximum variability of experimentally-feasible flux values across growth conditions. Adjusting $\gamma = 3.5 \pm 0.1$ resulted in a flux change of approximately $\pm$ 0.0078 (averaged across all conditions).

1.3. Flux Constraints

The uptake rate (i.e. lower bound) of carbon dioxide was fixed at -10 under all growth conditions, except for the low CO_2 growth condition (-0.01). In order to establish a protocol for specifying the variation in light uptake across growth conditions, a photon uptake rate (P_U) was calculated for each growth condition using a method similar to Vu *et al.* (Vu et al., 2012). The available light consumption (LC) under each condition (mmol) was multiplied by the surface area (SA) of the culture exposed to the light source (m^2), then the product was divided by the total available dry cell weight (DCW) of the culture (g per volume):

$$P_U = \frac{LC \times SA}{DCW} \tag{3}$$

ID	Condition	Specifics	Ref.
1	Dark oxic	Incubated in darkness prior to harvest, sparged in N_2	(Ludwig and Bryant, 2011)
2	Dark anoxic	Incubated in darkness prior to harvest	(Ludwig and Bryant, 2011)
3	High light	Illuminated at 900 μmol photons m^{-2} s^{-1} prior to harvest	(Ludwig and Bryant, 2011)
4	OD 0.4	Harvested at OD 730nm = 0.4	(Ludwig and Bryant, 2011)
5	OD 1.0	Harvested at OD 730nm = 1.0	(Ludwig and Bryant, 2011)
6	OD 3.0	Harvested at OD 730nm = 3.0	(Ludwig and Bryant, 2011)
7	OD 5.0	Harvested at OD 730nm = 5.0	(Ludwig and Bryant, 2011)
8	Low O_2	Sparged in N_2	(Ludwig and Bryant, 2011)
9	Low CO_2	Sparged with air [0.035% (v/v) CO_2]	(Ludwig and Bryant, 2012a)
10	N-limited	Cells washed in medium A (lacking NO_3^-) and resuspended	(Ludwig and Bryant, 2012a)
11	S-limited	Cells washed with $MgCl_2$	(Ludwig and Bryant, 2012a)
12	PO_4^{3-} limited	Cells washed w/o (PO_4^{3-}) harvested at OD = 0.7	(Ludwig and Bryant, 2012a)
13	Fe-limited	Cells washed in medium A with 720 μM deferoxamine mesylate B added at OD 0.35	(Ludwig and Bryant, 2012a)
14	NO_3^-	Standard growth in medium A (lacking $NaNO_3$) with 25 mM HEPES, 1 μM NiSO4, 12 mM $NaNO_3$	(Ludwig and Bryant, 2012a)
15	NH_3	Standard growth in medium A (lacking $NaNO_3$) with 25 mM HEPES, 1 μM $NiSO_4$ and 10 mM NH_4Cl	(Ludwig and Bryant, 2012a)
16	$CO(NH_2)_2$	Standard growth in medium A (lacking $NaNO_3$) with 25 mM HEPES, 1 μM $NiSO_4$ and 10 mM $CO(NH_2)_2$	(Ludwig and Bryant, 2012a)
17	Heat Shock	1h heat shock at 47°C	(Ludwig and Bryant, 2012b)
18	22°C	Standard growth at 22°C	(Ludwig and Bryant, 2012b)
19	30°C	Standard growth at 30°C	(Ludwig and Bryant, 2012b)
20	Oxidative stress	5 μM methyl viologen added 30 minutes prior to harvesting	(Ludwig and Bryant, 2012b)
21	Mixotrophic	Medium A+ supplemented with 10 mM glycerol	(Ludwig and Bryant, 2012b)
22	Low salt	Medium A+ containing 3 mM NaCl and 0.08 mM KCl	(Ludwig and Bryant, 2012b)
23	High salt	Medium A+ containing 1.5 M NaCl and 40 mM KCl	(Ludwig and Bryant, 2012b)
24	Standard control	Medium A+ at 38°C (*see caption for full details*)	(Ludwig and Bryant, 2012b)

Table S1: **Growth and stress conditions for *Synechococcus* sp. PCC 7002, Related to Figure 2.** The standard control condition was defined as: Medium A+ at 38°C, illuminated at 250 μmol photons $m^{-2}s^{-1}$, sparged in air with 1% (v/v) CO_2, with cells harvested at OD 730nm = 0.7.

As per the culture conditions described by Ludwig and Bryant (Ludwig and Bryant, 2011), the surface area of the culture exposed to the light source was calculated using the diameter of the culture tube (20mm) and the volume of the culture medium (25ml). The DCW of marine *Synechococcus* has been approximated at 0.35 gDW/L for an optical density (OD) of 1.0 at 750nm (Myers et al., 2013), which is close to other DCW estimates for *Synechococci* (Aikawa et al., 2014; Qiao et al., 2018). Given this estimate, a pre-established linear calibration for *Synechococcus* PCC 7942 cultures was used to calculate DCW for optical densities above OD 1.8 (i.e. OD 3.0 and 5.0) (Kato et al., 2017). In the plot (y = DCW, x = OD), since the same line could not be extrapolated below OD = 1.8 due to yielding negative DCW values, we adopted a piecewise linear approximation, and a separate linear equation was calculated for OD 0.4 and 0.7 between the lowest OD point (1.8,0.375) and the origin (0,0). The full specification of constraints for each growth condition and further information on their calculation is available in Supplementary Data 3.

1.4. Data Normalization

When creating the combined transcript-and-flux dataset, the two types of data (gene transcripts and reaction fluxes) were combined by converting them into fold change values. The fold changes of the RNA-Seq data was obtained by dividing the RPKM values for each growth condition by the average expression of three standard control replicates for that gene, as detailed in Section 1.2. In order to scale flux rates in the same range as the gene transcripts, we converted flux rates into fold change values by dividing the flux rates under these conditions by the flux rate derived for the standard condition. This provided a ratio of reaction activity between experimental conditions and the standard control. In order to account for the FBA solver tolerance, all the negligible flux rates (rate $< 10^{-4}$) were set to zero. When both experimental and control flux rates were negligible, fold change values were set to 1, while fold changes for divisions resulting in infinite values (due to the control flux being zero) were set to the maximum fold change for that dataset.

1.5. Principal Component Analysis (PCA)

Ascertaining the contribution of each condition to the construction of each dimension allowed us to detect conditions that deviated from the usual patterns, as well as those that were the greatest contributors to variance in the dataset. The PCA was conducted using the FactoMineR package in R (Lê et al., 2008), where conditions were described by gene transcripts and/or reaction fluxes (quantitative variables) for each pair of objectives, and the contribution of each condition to the dimensions and variance in the datasets was recorded.

Figure 2(a-c) displays PCA individual factor maps for the three objective pairs, showing the principal component scores of 24 individuals (which in our case are simulated growth conditions) described by 742 fluxes on the first two principal components. Figure 3(a), (c), (e) and (g) display individual factor maps for the three objective pairs and the respective principal component scores of 24 individuals described by 3187 transcripts (a), or 3929 multi-omic variables (c), (e) and (g), which include both transcripts and fluxes.

1.6. K-means Clustering

Gene transcripts, flux rates, or a combination of both measurements were used as variables for clustering growth conditions. To specify the appropriate number of clusters for running the k-means algorithm, we performed a silhouette analysis to assess the accuracy of cluster assignment by measuring the cohesion of data points within each cluster (given by a silhouette value for each variable). Both silhouette analysis and clustering were performed in MATLAB with the *silhouette* and *kmeans* functions, using the number of clusters that returned the highest silhouette values for the majority of points ($k = 6$).

Whilst computing correlation between genes across the profiles, the *zscore* function was used to standardize each of the profiles to have zero mean and unit variance. The pattern of clustering using the "cityblock" distance metric was compared between: (i) transcripts only (ii), fluxes only, (iii) both gene transcripts and fluxes. Multidimensional scaling was performed using a robust variation of the *mdscale* function in Matlab to circumvent colocation of points by multiplying dissimilarities by a scalar value (minimizing the squared stress criterion with 500 iterations of the iterative algorithm). The results of this clustering are reported in Figure 2(d-f) and Figure 3(b), (d), (f) and (h).

1.7. LASSO Regression

To relate the calculated flux distributions to *in-vivo* growth in our model, x is the data matrix consisting of input variables in the form of (i) gene transcripts (ii) flux rates, or (iii) both gene transcripts and flux rates, and y is the vector of growth rates in Table S2. Of the 23 transcriptomic profiles procured from the original studies, only 12 growth conditions had available (i) specified growth rates, (ii) specified doubling times, or (iii) standard growth curves (Ludwig and Bryant, 2011, 2012a,b). For these growth curves, the gradient between OD 0.4 and OD 0.7 was calculated as the growth rate, and all the other growth rates and doubling times were calculated relative to the standard growth rate. Although growing photoautotrophically in nature, heterotrophically-grown microalgae sometimes have higher growth rates, in the absence

of light shading and more space available to achieve greater cellular density (Kim et al., 2016).

Condition	Doubling time	Growth rate
Standard	100.000	0.075
N-limited	162.500	0.046
S-limited	150.000	0.050
P-limited	212.500	0.035
Nitrate	43.281	0.173
Ammonia	28.133	0.267
Urea	28.133	0.267
22°C	194.000	0.039
30°C	109.000	0.069
Mixotrophic	84.000	0.089
Low salt	98.000	0.077
High salt	270.000	0.028

Table S2: **Growth rates for LASSO and correlation analysis, Related to Figures 5 and 6.** Growth rates of 12 conditions calculated from growth curves previously published for *Synechococcus* sp. PCC 7002 (Ludwig and Bryant, 2011, 2012a,b). Doubling time is given as a fraction of standard conditions (where 100 is the standard condition).

The formula for the LASSO algorithm is specified in Eq. 4. We solved the regression problem in the form:

$$\min_{\beta_0,\beta}\left(\frac{1}{2N}\sum_{i=1}^{N}(y_i-\beta_0-x_i^T\beta)^2+\lambda\sum_{j=1}^{p}|\beta_j|\right), \quad (4)$$

where N is the number of conditions, y_i is the growth rate at the (i)th condition, p is the number of predictors indexed by j, x_i is a vector of p values at condition i, λ is a positive regularization parameter and β_0 and β represent the scalar and p-vector coefficients.

Our LASSO analysis reduced the list of recursive features by retaining positive and negative non-zero coefficients greater than 0.01. Furthermore, to provide a measure of predictive accuracy, we calculated the mean squared error (MSE) for all the fitted coefficients yielded as a result of the LASSO regression in each dataset. These values are recorded in Supplementary Data 2. The errors were at least one magnitude below the observed growth rates (in the response variable y), showing low deviation/variance between predicted and actual values and therefore good predictive accuracy overall.

1.8. Correlation Analysis

To find the strength of the association between growth rates and gene expression or flux values, the Pearson correlation coefficient was calculated between each vector of gene transcripts/flux rates and growth rates across conditions (see Table S2). For this analysis, all flux fold changes were converted into absolute (non-negative) values before calculating the correlation, therefore considering reversible reaction fluxes in absolute value to represent the activity of that reaction. For each gene/reaction, the mean predictor coefficient (MPC) was calculated by averaging across coefficients in all vectors for that predictor. The highest positively/negatively correlated genes or reactions for each dataset are plotted in Figure 5.

2. Supplemental Results

2.1. Clustering

For Biomass-ATP maintenance fluxes (Figure 2d), the clusters formed were: (i) dark oxic, dark anoxic, low O_2, low CO_2, nitrogen limitation, sulfur limitation, nitrate, ammonia, heat shock and mixotrophic; (ii) high light intensity, phosphate limitation, and standard control; (iii) OD 0.4, OD 3.0, 30°C and oxidative stress; (iv) OD 1.0 and urea; (v) OD 5.0, iron limitation and high salinity; (vi) 22°C and low salinity.

For Biomass-Photosystem I fluxes (Figure 2e), the clusters were: (i) dark oxic, dark anoxic, OD 3.0, OD 5.0, phosphate limitation, heat shock, mixotrophic and low salinity; (ii) high light intensity; (iii) OD 0.4; (iv) OD 1.0, nitrogen limitation and sulfur limitation; (v) low O_2 (vi) low CO_2, iron limitation, nitrate, ammonia, urea, 22°C, 30°C, oxidative stress, high salinity, and standard control.

For Biomass-Photosystem II fluxes (Figure 2f), the clusters were: (i) dark oxic, dark anoxic, OD 3.0, OD 5.0, phosphate limitation, heat shock, mixotrophic and low salinity; (ii) high light intensity; (iii) OD 0.4; (iv) OD 1.0, nitrogen limitation and sulfur limitation; (v) low O_2 (vi) low CO_2, iron limitation, nitrate, ammonia, urea, 22°C, 30°C, oxidative stress, high salinity, and standard control.

For the gene transcripts (plots a-b in Figure 3) the clusters formed were: (i) dark oxic, high light intensity, OD 0.4, OD 1.0, OD 3.0, OD 5.0, low O_2, nitrogen limitation, sulfur limitation, phosphate limitation, nitrate, ammonia, urea and low salinity; (ii) dark anoxic; (iii) low CO_2; (iv) iron limitation; (v) heat shock; (vi) 22°C, 30°C, oxidative stress, high salinity, and standard control.

When combining both transcript and flux data (Figure 3 c-h), the clusters formed were less distinct. For the Biomass-ATP objective pair, the clusters were: (i) dark oxic, OD 1.0, OD 3.0, low CO_2, nitrogen limitation, sulfur limitation, iron limitation, nitrate, ammonia, urea, 30°C, oxidative stress and high salinity; (ii) dark anoxic, low O_2, phosphate limitation, heat shock, mixotrophic, and standard control; (iii) high light intensity; (iv) OD 0.4 (v) OD 5.0 (vi) 22°C and low salinity.

For the Biomass-PI objective pair, the clusters were: (i) dark oxic, OD 1.0, OD 3.0, OD 5.0, low O_2, low CO_2, nitrogen limitation, sulfur limitation, nitrate,

ammonia, urea, heat shock, 22°C, 30°C, oxidative stress, high salinity, and standard control; (ii) dark anoxic; (iii) high light intensity and OD 0.4; (iv) phosphate limitation and low salinity; (v) iron limitation; (vi) mixotrophic.

For the Biomass-PII objective pair, the clusters were: (i) dark oxic, OD 1.0, OD 3.0, OD 5.0, low O_2, low CO_2, nitrogen limitation, iron limitation, nitrate, ammonia, urea, heat shock, 22°C, 30°C, oxidative stress, high salinity, and standard control; (ii) dark anoxic; (iii) high light intensity and OD 0.4; (iv) sulfur limitation; (v) phosphate limitation; (vi) mixotrophic and low salinity.

2.2. Principal Component Contributions to Variance

Table S3 displays contributions of the fluxes to the principal components in the form of a list of the top ten reactions contributing to variation in the datasets for all three objective pairs. Among the highest contributors to variance for all objectives was inorganic diphosphate (IODP), which catalyzes the hydrolysis of phosphorous acid anhydrides. Additionally, aspartate transaminase (ASPTA1) is involved in the transfer of an amino group from aspartate to alpha-ketoglutarate during which glutamate and oxaloacetate are formed. In the absence of sufficient oxygen, photoautotrophs such as *Synechococcus* are capable of switching from aerobic respiration to anaerobic fermentation in order to produce ATP for cellular metabolic processes. For all objective pairs, formate exchange (EX_FOR_E) and transport (FORT) were among the highest contributors to variance. Methenyltetrahydrofolate cyclohydrolase (MTHFC) and methylenetetrahydrofolate dehydrogenase (MTHFD) are intermediates in folate biosynthesis that are involved in the Wood-Ljungdahl pathway (aka reductive acetyl coenzyme A pathway), where carbon fixation by acetyl-CoA synthase and fermentative respiration occur concurrently and acetate is generated as the end product((Woo and Jang, 2019)). The exchange and transport of formate (EX_FOR_E and FORT) are also linked with folate biosynthesis, since formate can be converted into tetrahydrofolate. Transketolase (TK1 and TK2) as well as sedoheptulose bisphosphatase are also utilized in the pentose phosphate pathway. The pentose phosphate pathway is an important source of NADPH, a reducing agent used to drive the numerous oxidation-reduction reactions throughout the central carbon metabolic pathways. In *Synechocystis* sp. PCC 6803, higher metabolic flux was observed for NADPH production under low light conditions since the oxidative pentose phosphate pathway provides an alternative route for NADPH production (Ueda et al., 2018). For the biomass-photosystem I objective pair, glutamate exchange (EX_GLU_E) and transport of L-glutamate(GLUSYM) and sodium (NAT3) were represented in the highest contributors. For the biomass-photosystem II objective pair, the key respiratory enzymes glucose-6-isomerase (PGI) and pyruvate dehydrogenase (PDH) as well as ribulose 5-phosphate 3-epimerase (RPE) from the pentose phosphate pathway were included. The pentose phosphate pathway runs parallel to glycolysis and is mainly responsible for the synthesis of amino acid and nucleotide precursors. Pyruvate dehydrogenase is among the most important enzymes in central carbon metabolism since it is responsible for producing acetyl-CoA.

2.3. LASSO Regression

Reducing the number of predictors in a regression model enables identification of important predictors, elimination of redundant predictions, and generation of shrinkage estimates with lower predictive errors than ordinary least squares regression. LASSO employs L1 regularization, which penalizes the sum of absolute values of all the coefficients; this sets the coefficients of unnecessary or recursive features equal to zero, resulting in a sparser matrix. The formula for the algorithm is specified in Eq. 4 in the main text. Tables S6, S7, and 2.3 list all non-zero coefficients and the transcripts/fluxes they are associated with. For each gene/reaction, the mean predictor coefficient (MPC) is calculated by averaging across coefficients in all vectors for that predictor.

When applying the LASSO algorithm to the flux-only dataset (Table S7), the coefficients identified primarily belonged to pathways associated with the metabolism of nucleotides, co-factors and vitamins, but also those that converge in order to fulfil a common objective i.e. energy generation from carbohydrate and lipid metabolism, or the exchange and transport of specific metal ions. Of particular interest is cobalt transport (COBALTT5), since this reaction yielded one of the highest coefficient values. Many cyanobacterial species (including *Synechococcus*) utilize nickel, copper, zinc, and cobalt-containing enzymes as protein co-factors that constitute important components of the photosynthetic machinery (Huertas et al., 2014). Furthermore, these co-factors contribute to mechanisms for survival in iron-depleted environments (Palenik et al., 2003), such as the secretion of siderophores to chelate iron. In the form of the photosystem II manganese-stabilizing polypeptide (psbO), manganese plays a critical role in photosystem II function by catalyzing the light-induced dissociation of water to molecular oxygen (Bartsevich and Pakrasi, 1995; Liu et al., 2018). Inactivation of the manganese transport system (MNABC) leads to an invariable loss of activity in photosystem II, consequently affecting the photosynthetic process as a whole (Shcolnick and Keren, 2006). Similarly, the molybdenum exchange and transport reactions (EX_MOBD_E and MOBDABC) represent the

Biomass - ATP maintenance				
No.	**Rxn**	Reaction Name	Dim 1	Dim 2
1	IODP	inorganic diphosphatase	0.307	0.012
2	TKT1	transketolase	0.307	0.016
3	NDPK1	nucleoside-diphosphate kinase (ATP:GDP)	0.307	0.016
4	RPE	ribulose 5-phosphate 3-epimerase	0.307	0.016
5	TKT2	transketolase	0.307	0.016
6	FBA3	Sedoheptulose 1,7-bisphosphate D-glyceraldehyde-3-phosphate-lyase	0.307	0.016
7	SBP	sedoheptulose-bisphosphatase	0.307	0.016
8	FBA	fructose-bisphosphate aldolase	0.307	0.016
9	FBP	fructose-bisphosphatase	0.307	0.016
10	NDPK3	nucleoside-diphosphate kinase (ATP:CDP)	0.307	0.016
Biomass - Photosystem I				
No.	**Rxn**	Reaction Name	Dim 1	Dim 2
1	ASPTA1	aspartate transaminase	0.270	0.001
2	GLUSYM	L-Glutamate transport in via sodium symport	0.270	0.004
3	EX_GLU_E	glutamate exchange	0.270	0.004
4	NAT3	sodium transport out via proton antiport	0.270	0.004
5	IODP	inorganic diphosphatase	0.270	0.004
6	EX_FOR_E	formate exchange	0.270	0.003
7	FORT	formate transport via diffusion	0.269	0.003
8	MTHFC	methenyltetrahydrofolate cyclohydrolase	0.269	0.004
9	MTHFD	methylenetetrahydrofolate dehydrogenase (NADP)	0.269	0.004
10	FTHFD	formyltetrahydrofolate deformylase	0.269	0.003
Biomass - Photosystem II				
No.	**Rxn**	Reaction Name	Dim 1	Dim 2
1	ASPTA1	aspartate transaminase	0.272	1.36×10^{-5}
2	IODP	inorganic diphosphatase	0.272	0.001
3	PDH	pyruvate dehydrogenase	0.272	0.001
4	MTHFC	methenyltetrahydrofolate cyclohydrolase	0.272	4.49×10^{-4}
5	MTHFD	methylenetetrahydrofolate dehydrogenase (NADP)	0.272	4.49×10^{-4}
6	TKT1	transketolase	0.272	3.57×10^{-4}
7	EX_FOR_E	formate exchange	0.272	1.23×10^{-4}
8	FORT	formate transport via diffusion	0.272	0.001
9	PGI	glucose-6-phosphate isomerase	0.272	0.001
10	RPE	ribulose 5-phosphate 3-epimerase	0.272	0.001

Table S3: **Contributions of reactions to variance, Related to Figure 2.** Top ten contributions of individual reaction fluxes to principal component variables for the first two dimensions when PCA is performed on the fluxes for three pairs of objectives (biomass-ATP maintenance, biomass-photosystem I, biomass-photosystem II).

No.	**Gene**	**COG category**	**CY category**	**CY subcategory**	**Dim 1**	**Dim 2**
1	A0445	AA transport and metabolism	Transport and binding proteins	NA	0.125	0.002
2	A1216	Translation, ribosomal structure and biogenesis	Translation	Aminoacyl tRNA synthetases and tRNA modification	0.124	5.16×10^{-5}
3	A0327	Energy production and conversion	Hypothetical	NA	0.124	0.007
4	A1341	None	Unknown	NA	0.123	0.0006
5	A1274	AA transport and metabolism/Cell wall, membrane, envelope biogenesis	AA biosynthesis	Aspartate family	0.121	0.0008
6	A0831	Coenzyme transport and metabolism	Biosynthesis of cofactors, prosthetic groups, and carriers	Thiamin	0.119	8.66×10^{-5}
7	A1173	Secondary metabolites biosynthesis, transport and metabolism	Fatty acid, phospholipid and sterol metabolism	NA	0.119	0.0003
8	A2554	Carbohydrate transport and metabolism	Transport and binding proteins	NA	0.111	7.59×10^{-8}
9	A0939	Signal transduction mechanisms	Hypothetical	NA	0.111	0.011
10	A1414	Energy production and conversion	Energy Metabolism	Pyruvate and acetyl-CoA metabolism	0.110	0.002

Table S4: **Contributions of genes to variance, Related to Figure 3.** Top ten contributions of genes to principal component variables for the first two dimensions when PCA is performed on the gene transcripts. COG category refers to the Cluster of Orthologous Groups that the gene belongs to, whilst CY category is its functional category according to Cyanobase (Fujisawa et al., 2016).

Biomass - ATP maintenance						
No.	Gene/Rxn	COG category/Rxn description	CY category/subsystem	CY subcategory	Dim 1	Dim 2
1	A0831	Coenzyme transport and metabolism	Biosynthesis of cofactors, prosthetic groups, and carriers	Thiamin	0.118	3.31×10^{-5}
2	A1274	AA transport and metabolism/Cell wall, membrane, envelope biogenesis	Amino acid biosynthesis	Aspartate family	0.118	0.002
3	A1216	Translation, ribosomal structure and biogenesis	Translation	Aminoacyl tRNA synthetases and tRNA modification	0.118	0.002
4	A0445	AA transport and metabolism	Transport and binding proteins	NA	0.117	0.002
5	A1341	None	Unknown	NA	0.116	0.005
6	A0327	Energy production and conversion	Hypothetical	NA	0.115	0.012
7	A1173	Secondary metabolites biosynthesis, transport and metabolism	Fatty acid, phospholipid and sterol metabolism	NA	0.115	0.003
8	A2796	Replication, recombination, and repair	DNA replication, restriction, modification, recombination, and repair	NA	0.110	0.003
9	A0743	Lipid transport and metabolism	Other	Other	0.107	0.003
10	A1414	Energy production and conversion	Energy Metabolism	Pyruvate and acetyl-CoA metabolism	0.107	0.003
Biomass - Photosystem I						
No.	Gene/Rxn	COG category/Rxn description	CY category/subsystem	CY subcategory	Dim 1	Dim 2
1	A0831	Coenzyme transport and metabolism	Biosynthesis of cofactors, prosthetic groups, and carriers	Thiamin	0.119	0.001
2	A1274	AA transport and metabolism/Cell wall, membrane, envelope biogenesis	Amino acid biosynthesis	Aspartate family	0.115	0.005
3	A0445	AA transport and metabolism	Transport and binding proteins	NA	0.114	0.006
4	A1216	Translation, ribosomal structure and biogenesis	Translation	Aminoacyl tRNA synthetases and tRNA modification	0.114	0.007
5	A1341	None	Unknown	NA	0.112	0.009
6	A1173	Secondary metabolites biosynthesis, transport and metabolism	Fatty acid, phospholipid and sterol metabolism	NA	0.111	0.005
7	A0327	Energy production and conversion	Hypothetical	NA	0.111	0.016
8	A2796	Replication, recombination, and repair	DNA replication, restriction, modification, recombination, and repair	NA	0.110	0.001
9	A0743	Lipid transport and metabolism	Other	Others	0.107	2.69×10^{-4}
10	A1405	Signal transduction mechanisms	Other	Drug and analog sensitivity	0.105	0.001
Biomass - Photosystem II						
No.	Gene/Rxn	COG category/Rxn description	CY category/subsystem	CY subcategory	Dim 1	Dim 2
1	A0831	Coenzyme transport and metabolism	Biosynthesis of cofactors, prosthetic groups, and carriers	Thiamin	0.119	7.48×10^{-4}
2	A1274	AA transport and metabolism/Cell wall, membrane, envelope biogenesis	Amino acid biosynthesis	Aspartate family	0.115	0.005
3	A0445	AA transport and metabolism	Transport and binding proteins	NA	0.115	0.006
4	A1216	Translation, ribosomal structure and biogenesis	Translation	Aminoacyl tRNA synthetases and tRNA modification	0.114	0.007
5	A1341	None	Unknown	NA	0.113	0.009
6	A1173	Secondary metabolites biosynthesis, transport and metabolism	Fatty acid, phospholipid and sterol metabolism	NA	0.112	0.005
7	A0327	Energy production and conversion	Hypothetical	NA	0.111	0.016
8	A2796	Replication, recombination, and repair	DNA replication, restriction, modification, recombination, and repair	NA	0.111	0.001
9	A0743	Lipid transport and metabolism	Other	Other	0.108	2.57×10^{-4}
10	A1405	Signal transduction mechanisms	Other	Drug and analog sensitivity	0.106	0.001

Table S5: **Contributions of genes and reactions to variance, Related to Figure 3.** Top ten contributions of gene transcripts and reaction fluxes to principal component variables for the first two dimensions when PCA is performed on the gene transcripts and fluxes for three pairs of objectives (biomass-ATP maintenance, biomass-photosystem I, biomass-photosystem II).

uptake and transport of molybdenum by cyanobacterial ABC transporters to form cofactors (e.g. Moco or FeMoco) that assist in nitrogen fixation (Shvarev and Maldener, 2018; Demtröder et al., 2019). GMPS2 (guanosine monophosphate synthetase) was among the top negative LASSO coefficients and catalyzes the reversible conversion between xanthosine 5' phosphate and GMP (guanosine monophosphate) and also between L-glutamine and L-glutamate. Significantly, the nitrate reduction reaction (FDNOR1) was one of the main negative LASSO coefficients retained in the Biomass-Photosystem I flux data. Ferredoxins are iron and sulfur-containing proteins that act as electron carriers during many important metabolic processes including oxidation-reduction, photosynthesis, and nitrogen fixation (Lea-Smith et al., 2016). More specifically, nitrate ferredoxin acts as a significant electron sink during the reversible, redox inter-conversion of nitrite to nitrate during photosynthesis (Flores et al., 2005; Qian et al., 2016).

The positive coefficients for reactions previously identified during PCA were also retained for the Biomass-Photosystem I objective pair. ASPO5 (anaerobic L-aspartate oxidase) was identified as the highest positive LASSO coefficient in the ATP objective pair. This anaerobic reaction uses fumarate as a terminal electron acceptor instead of oxygen, producing succinate.

Succinate dehydrogenase (SUCD1Itlm and SUCD1Icpm) catalyzes the interconversion of succinate to fumarate as a product of NADH oxidation in the TCA cycle, a process which can still occur in *Synechococcus* sp. PCC 7002 during fermentative metabolism in dark, anoxic conditions (McNeely et al., 2010). In a diurnal GSMM accounting for fluctuations in light availability in the phototrophic metabolism of *Synechocystis* sp. PCC 6803, an increased flux towards succinate was observed during the dark period, driving ATP production through respiratory electron transport (Sarkar et al., 2019). Other correlated reactions were associated with the oxidation, transport and exchange of D-lactate and water, synthesis of intermediates used in central carbon metabolism, and for synthesis of intermediates formed during amino acid biosynthesis and catabolism. The negatively correlated reactions included dehydrogenases in the TCA cycle that catalyze the reversible oxidation of carbohydrates or amino acids (L-aspartate, L-alanine, D-lactate) into pyruvate or oxaloacetate, which can be utilized in the central metabolic pathways (glycolysis or the TCA cycle).

2.4. Pearson Correlation Analysis

Similarly to LASSO, the Pearson correlation coefficient was calculated between each dataset of gene transcripts/flux rates (x) and growth rates (y) listed in Table S2 in the main text. For this analysis, all transcript and flux fold changes were converted into absolute (non-negative) values. Plots for the top ten positively/negatively correlated genes are given in Figure S1. Tables S9, S10, S11, S12, and 2.4 provide additional information on the function of each gene or reaction as well as its Pearson correlation coefficient. Additionally, Figure 6 (c) displays the mean absolute PCC values according to subsystems defined in the GSMM.

For all three objective pairs, genes involved in succinate dehydrogenation (SUCD1Itlm/SUCD1Icpm), efflux (SUCCt2b) or exchange (EX_succ_E) were found to be positively correlated with growth. In addition to succinate dehydrogenase, fumarase (FUMH), succinate semialdehyde dehydrogenase (SSALY) and 2-oxoglutarate (2OGDC) were positively correlated with growth in Biomass-ATP maintenance. All three of these enzymes play an important role in the interconversion of compounds within the TCA cycle. Overall, there were a high number of reactions that yielded similar positive coefficients for Biomass - Photosystem I and Biomass - Photosystem II data (Table 2.4). These reactions belonged to pathways relating to amino acid, carbohydrate, exchange and transport metabolism. NADPH dehydrogenases are among the top negative coefficients for all three flux datasets. Cyanobacteria consume a large amount of the cofactors NADPH and NADH whilst reducing nitrate to catabolize glycogen under dark anaerobic conditions (Qian et al., 2016). NADH dehydrogenase (type II) is a protein that catalyzes the electron transfer from NADH to a quinone molecule via a flavin co-factor (Heikal et al., 2014). In the case of *Synechococcus* and other cyanobacteria, this molecule is plastoquinone, which plays a critical role as a mobile electron carrier during the light-dependent reactions of photosynthesis. NADH type II is formed when plastoquinone combines with NADH to form NADH:plastoquinone oxidoreductase, which acts as a single-subunit flavoenzyme. In photosystem II, plastoquinone is doubly reduced to plastoquinol, acting as a terminal electron acceptor (McConnell et al., 2011). The reduced plastoquinone (plastiquinol) pool serves as an electron buffer, maintaining a consistent charge between photosystem II and photosystem I (Peltier et al., 2016). The reduction from plastiquinone to plastiquinol can occur both in the cytoplasm and the thylakoid membrane, either with or without a proton pump. Reactions for water exchange (EX_H20_E) and transport (H20T5) have the highest negative correlation with growth in the Biomass-ATP maintenance flux data. In addition to the photosystem II reaction (PSIIR), water and oxygen exchanged between the thylakoid lumen and cytosol are among the top negative coefficients in the Biomass-Photosystem II flux data. In cyanobacteria, water plays an important role in photo-

Gene	COG category	CY category	CY subcategory	MPC
A0639	None	Photosynthesis and respiration	Phycobilisome	0.036
A0575	None	Hypothetical	NA	0.012
A0720	Posttranslational modification, protein turnover, chaperones	Central intermediary metabolism	Others	0.011
A0083	General	Translation	Degradation of proteins peptides and glycopeptides	-0.012
A1376	Cell cycle control, cell division, chromosome partitioning	Hypothetical	NA	-0.025
G0060	Inorganic ion transport and metabolism	Transport and binding proteins	NA	-0.049

Table S6: **Non-zero coefficients (>0.01) retained by LASSO regularizer for all gene transcripts, Related to Section 3.4 LASSO regression and Figure 5.** The mean predictor coefficient (MPC) is calculated by averaging across coefficients in all vectors for that predictor. Genes yielding positive coefficients are associated with photosynthesis, respiration, and protein modification. Genes yielding negative predictor coefficients are involved in protein degradation, cell division, and transport of inorganic ions.

Biomass - ATP maintenance			
Reaction	Description	Subsystem	MPC
ASPO5	L-aspartate oxidase	AA Metabolism	11.253
GLNS	glutamine synthetase	AA Carbohydrate and Energy Metabolism	2.387
ILEABC	L-isoleucine transport via ABC system	Extracellular Transport	0.247
LEUABC	L-leucine transport via ABC system	Extracellular Transport	0.063
GMPS2	GMP synthase (glutamine-hydrolysing)	Nucleotide Metabolism	-3.589
IMPD	IMP dehydrogenase	Nucleotide Metabolism	-0.192
Biomass - Photosystem I			
Reaction	Description	Subsystem	MPC
COBALTT5	cobalt transport in/out via permease (no H+)	Transport	267.011
GLUSZ	glutamate synthase (Ferredoxin)	Energy and Carbohydrate Metabolism	88.057
ADSL2R	adenylosuccinate lyase	Nucleotide Metabolism and AA Metabolism	45.372
ATPM	ATP maintenance requirement	Nucleotide Metabolism	41.438
NTD7	5"-nucleotidase (AMP)	Nucleotide Metabolism	38.909
PRAGS	phosphoribosylglycinamide synthetase	Nucleotide Metabolism	14.157
ADCL	4-aminobenzoate synthase	Metabolism of cofactors and vitamins	13.206
THFAT	tetrahydrofolate aminomethyltransferase	Metabolism of cofactors and vitamins	12.862
PIABC	phosphate transport via ABC system	Transport	11.172
EX_GLYC_E	glycerol exchange	Exchange Reaction	9.735
ME2	malic enzyme (NADP)	Carbohydrate Metabolism and Energy Metabolism	-176.122
MNABC	manganese transport via ABC system	Transport	-3.832
FDNOR1	ferredoxin-NADP reductase	Energy Metabolism	-3.641
HTDHL6	(3R)-3-Hydroxypalmitoyl-[acyl-carrier-protein] hydro-lyase	Lipid metabolism and Metabolism of cofactors and vitamins	-2.140
HOXPRX	2-hydroxy-3-oxopropionate reductase (NAD)	Carbohydrate Metabolism	-1.509
BTMAT1	Butyryl-[acyl-carrier protein]:malonyl-CoA C-acyltransferase	Metabolism of cofactors and vitamins	-1.098
GTPCI	GTP cyclohydrolase I	Metabolism of cofactors and vitamins	-0.950
EX_COBALT2_E	cobalt exchange	Exchange Reaction	-0.913
DB4PS	3,4-Dihydroxy-2-butanone-4-phosphate	Metabolism of cofactors and vitamins	-0.821
GAPD_NADP	glyceraldehyde-3-phosphate dehydrogenase (NADP) (phosphorylating)	Energy Metabolism	-0.817
Biomass - Photosystem II			
Reaction	Description	Subsystem	MPC
ADCL	4-aminobenzoate synthase	Metabolism of cofactors and vitamins	367.805
EX_GLYC_E	glycerol exchange	Exchange Reaction	119.944
G3PD2	glycerol-3-phosphate dehydrogenase (NADP)	Lipid metabolism	61.787
ALALIG	D-alanine-D-alanine ligase (reversible)	Cell wall and Metabolism of other amino acids	48.533
EX_MN2_E	manganese exchange	Exchange Reaction	4.757
EX_PTRC_E	putrescine exchange	Exchange Reaction	4.011
PTRCABC	putrescine transport via ABC system	Extracellular Transport	0.301
GMPS2	GMP synthase (glutamine-hydrolysing)	Nucleotide Metabolism	-0.829

Table S7: **Non-zero coefficients (>0.01) retained by LASSO regularizer for fluxes in the three pairs of objectives, Related to Section 3.4 LASSO regression and Figure 5.** Flux rates below 10^{-4} were not considered in order to account for solver tolerance during optimization. The coefficients retained are related to pathways involved in energy metabolism - fatty acid synthesis, transport/exchange, nucleotide metabolism, carbohydrate metabolism, and amino-acid metabolism.

Biomass - ATP maintenance				
Gene/Rxn	COG category/Rxn description	CY category/subsystem	CY subcategory	MPC
A0639	None	Photosynthesis and respiration	Phycobilisome	0.036
A0575	None	Hypothetical	NA	0.012
A0720	Posttranslational modification, protein turnover, chaperones	Central intermediary metabolism	Others	0.011
G0060	Inorganic ion transport and metabolism	Transport and binding proteins	NA	-0.049
A1376	Cell cycle control, cell division, chromosome partitioning	Hypothetical	NA	-0.025
A0083	General	Translation	Degradation of proteins peptides and glycopeptides	-0.012
Biomass - Photosystem I				
Gene/Rxn	COG category/Rxn description	CY category/subsystem	CY subcategory	MPC
A0639	None	Photosynthesis and respiration	Phycobilisome	0.036
A0575	None	Hypothetical	NA	0.012
A0720	Posttranslational modification, protein turnover, chaperones	Central intermediary metabolism	Others	0.011
G0060	Inorganic ion transport and metabolism	Transport and binding proteins	NA	-0.049
A1376	Cell cycle control, cell division, chromosome partitioning	Hypothetical	NA	-0.025
A0083	General	Translation	Degradation of proteins peptides and glycopeptides	-0.012
Biomass - Photosystem II				
Gene/Rxn	COG category/Rxn description	CY category/subsystem	CY subcategory	MPC
A0639	None	Photosynthesis and respiration	Phycobilisome	0.036
A0575	None	Hypothetical	NA	0.012
A0720	Posttranslational modification, protein turnover, chaperones	Central intermediary metabolism	Others	0.011
G0060	Inorganic ion transport and metabolism	Transport and binding proteins	NA	-0.049
A1376	Cell cycle control, cell division, chromosome partitioning	Hypothetical	NA	-0.025
A0083	General	Translation	Degradation of proteins peptides and glycopeptides	-0.012

Table S8: **Non-zero coefficients (>0.01) retained by LASSO regularizer for all features (gene transcripts and fluxes) with the three objective pairs, Related to Section 3.4 LASSO regression and Figure 5.** Only gene coefficients are retained, which are the same predictors as Table S6.

system II when it splits to produce molecular oxygen during the light-driven flux of protons from the cytoplasm to the luminal side of the thylakoid membrane, enabling a transmembrane electric potential for the diffusion of metal ions (Checchetto et al., 2012). Similar to results of the LASSO, the nitrate reduction reaction (FDNOR1) has the highest negative Pearson coefficient in the Biomass-Photosystem I flux data. The NAD-dependent malic enzyme (ME2) yielded a highly positive correlation coefficient for the Biomass- ATP maintenance data but a negative coefficient for the Biomass - Photosystem I data. Its function within the TCA cycle is the oxidative decarboxylation of malate and NAD+ to pyruvate and NADH.

When the gene transcripts were analyzed in isolation (Table S9), the highest Pearson correlation coefficients were given by (i) genes relating to photosynthesis and energy metabolism (A0639, A1008, A1802), and (ii) genes relating to amino acid biosynthesis and protein assembly (A2457, A0881). The phycobilisome is a large, light-harvesting membrane complex in photosystem II comprising chromophorylated phycobiliproteins and linker peptides (Boulay et al., 2008). Phycobilisomes assist cyanobacteria in adapting to varying light conditions by adjusting their size and structure to accommodate the flow of energy (Noreña-Caro and Benton, 2018). These are evidently important features to retain in the model for a phototrophic cyanobacterium as photophysiological measurements associated with photosystems tend to increase with the growth rate (Watanabe et al., 2014). Correspondingly, nutrient limitation causes degradation of the phycobilisomes, which activates mechanisms responsible for maintaining photosynthetic efficiency (Jackson et al., 2015; Clark et al., 2018; Saha et al., 2016).

When the gene expression and flux datasets were combined, all of the same genes are identified in Table 2.4 as in Table S9 and there was no difference in coefficients between the objectives (i.e. Biomass - ATP maintenance, Biomass - Photosystem I, Biomass - Photosystem II). This supports our idea of using fluxes separately as additional predictive omic features in further machine learning analyses (Zampieri et al., 2019; Yang et al., 2019; Culley et al., 2020; Zhang et al., 2020), for predicting growth rate or production rates for biotechnologically-relevant metabolites.

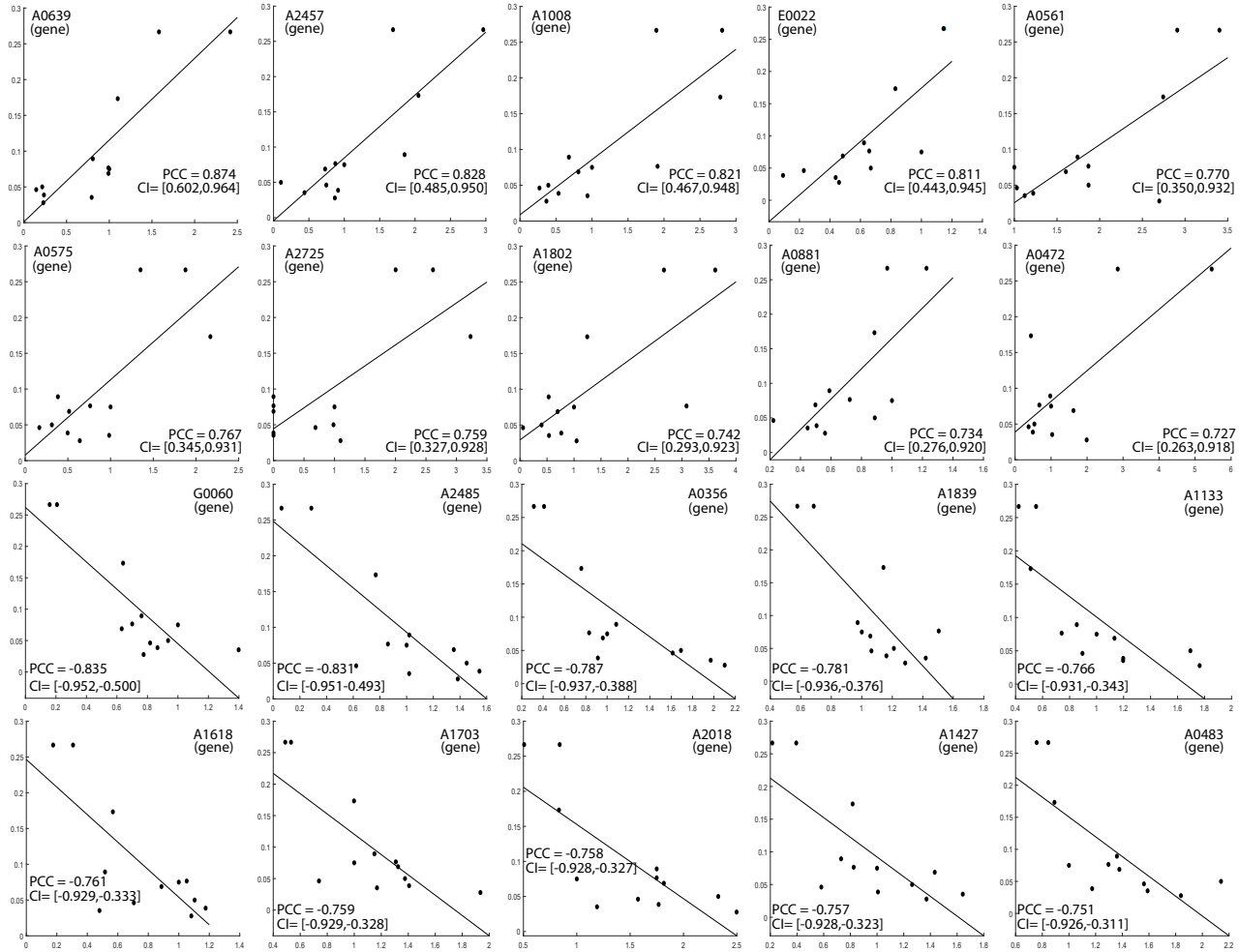


Figure S1: **Top ten Pearson correlation coefficients (PCC) with their respective 95% confidence intervals (CI) between gene transcript values (x) and growth rates (y), Related to Figure 5.** A list of each gene and its respective PCC is provided in Table S9.

Gene	COG category	CY category	CY subcategory	PCC
A0639	None	Photosynthesis and respiration	Phycobilisome	0.874
A2457	Post-translational modification, protein turnover, chaperones	Cellular processes	Chaperones	0.828
A1008	None	Photosynthesis and respiration	Photosystem I	0.821
E0022	None	Unknown	NA	0.811
A0561	None	Hypothetical	NA	0.770
A0575	None	Hypothetical	NA	0.767
A2725	None	NA	NA	0.759
A1802	Energy production and conversion/Secondary metabolites biosynthesis, transport and metabolism	Photosynthesis and respiration	CO_2 fixation	0.742
A0881	AA transport and metabolism	AA biosynthesis	Aspartate family	0.734
A0472	None	Hypothetical	NA	0.727
G0060	Inorganic ion transport and metabolism	Transport and binding proteins	NA	-0.835
A2485	General	Hypothetical	NA	-0.831
A0356	General	Translation	Degradation of proteins peptides and glycopeptides	-0.787
A1839	None	Other	Other	-0.781
A1133	Replication, recombination, and repair	Translation	Aminoacyl tRNA synthetases and tRNA modification	-0.766
A1618	AA transport and metabolism	Hypothetical	NA	-0.761
A1703	AA transport and metabolism	Energy Metabolism	Amino acids and amines	-0.759
A2018	General	Hypothetical	NA	-0.758
A1427	Coenzyme transport and metabolism	Biosynthesis of cofactors, prosthetic groups, and carriers	Pantothenate	-0.757
A0483	General	Hypothetical	NA	-0.751

Table S9: **Top ten Pearson correlation coefficients between gene transcripts and growth rates for 11 conditions, Related to Figures 5 and S1.** The highest correlation coefficients are given by genes relating to photosynthesis and energy metabolism or amino acid biosynthesis and protein assembly.

Biomass - ATP maintenance			
Reaction	Description	Subsystem	PCC
SUCD1Itlm	succinate dehydrogenase	Carbohydrate Metabolism	0.683
SUCD1Icpm	succinate dehydrogenase	Carbohydrate Metabolism	0.683
ME2	malic enzyme (NADP)	Carbohydrate Metabolism and Energy Metabolism	0.683
CA2T3	Ca2 transport via ion channels	Transport	0.605
GLNS	glutamine synthetase	AA Carbohydrate and Energy Metabolism	0.605
CA2T2	calcium transport out via proton antiport	Transport	0.605
EX_NH4_E	ammonia exchange	Exchange Reaction	0.536
NH4T	ammonium transport via diffusion	Transport	0.536
FUMH	fumarase	AA Metabolism	0.527
GLUSYM	L-Glutamate transport in via sodium symport	Transport	0.519
EX_GLU_E	glutamate exchange	Exchange Reaction	0.519
EX_H2O_E	water exchange	Exchange Reaction	-0.524
H2OT5	H2O transport via diffusion	Transport	-0.524
NADPHPQ9cpm	NADPH dehydrogenase (plastoquinone-9 and 4 protons)	Energy Metabolism	-0.391
ALAD_LR	L-alanine dehydrogenase (reversible)	AA Metabolism	-0.388
LDH_D	D-lactate dehydrogenase	Carbohydrate Metabolism	-0.381
lac_d2	D-lactate transport	Transport	-0.381
EX_lac_d_E	D-lactate exchange	Exchange Reaction	-0.381
NADPHPQ9tlm	NADPH dehydrogenase (plastoquinone-9 and 4 protons)	Energy Metabolism	-0.376
THD2	NAD(P) transhydrogenase	Metabolism of cofactors and vitamins	-0.353
ASPT	L-aspartase	AA Metabolism	-0.324

Table S10: **Top ten Pearson correlation coefficients between Biomass - ATP maintenance fluxes (x) and growth rates (y), Related to Figures 5 and S2.** The flux rates used to calculate PCC were absolute values of the fluxes calculated during bi-level regularized FBA. Reactions that yield the highest positive correlation coefficients belong to pathways relating to amino acid, carbohydrate, exchange and transport metabolism whereas reactions involved in nucleotide metabolism are negatively correlated with the growth rate.

Biomass - Photosystem I			
Reaction	Description	Subsystem	PCC
SUCCt2b	Succinate efflux via proton symport	Transport	0.378
EX_succ_E	Succinate exchange	Exchange	0.378
ASPT	L-aspartase	AA Metabolism	0.364
HDH2	Bidirectional Hydrogenase (NADP/NADPH)	Hydrogen Metabolism	0.355
NADPHPQ9tlm	NADPH dehydrogenase (plastoquinone-9 and 4 protons)	Energy Metabolism	0.349
NAT3	sodium transport out via proton antiport	Transport	0.343
GLUSYM	L-Glutamate transport in via sodium symport	Transport	0.343
EX_GLU_E	glutamate exchange	Exchange	0.343
SSALY	succinate-semialdehyde dehydrogenase (NADP)	AA Metabolism and Carbohydrate Metabolism	0.343
2OGDC	2-oxoglutarate decarboxylase	Exchange	0.343
FDNOR1	ferredoxin-NADP reductase	Energy Metabolism	-0.427
ME2	malic enzyme (NADP)	Carbohydrate Metabolism and Energy Metabolism	-0.259
GAPD_NADP	glyceraldehyde-3-phosphate dehydrogenase (NADP) (phosphorylating)	Energy Metabolism	-0.252
ACONT	aconitase	Carbohydrate Metabolism	-0.246
CS	citrate synthase	Carbohydrate Metabolism	-0.246
ICDHY	isocitrate dehydrogenase (NADP)	Metabolism of other amino acids and Energy Metabolism	-0.246
P5CD	1-pyrroline-5-carboxylate dehydrogenase	AA Metabolism	-0.242
FDPQ	Cyclic reaction (ferredoxin:plastoquinol)	Energy Metabolism	-0.240
CA2ABC1	calcium efflux via ABC system	Transport	-0.239
COABC	Cobalt transport via ABC system	Transport	-0.239
DADPEP	D-ala-D-ala dipeptidase	AA Metabolism	-0.239
DAGK_SYN	diacylglycerol kinase (Synechococcus)	Lipid metabolism	-0.239
EX_PI_E	phosphate exchange	Exchange Reaction	-0.239
FTHFCL	5-formyltetrahydrofolate cyclo-ligase	Metabolism of cofactors and vitamins	-0.239
NTD7	5"-nucleotidase (AMP)	Nucleotide Metabolism	-0.239
PIABC	phosphate transport via ABC system	Transport	-0.239
THFAT	tetrahydrofolate aminomethyltransferase	Metabolism of cofactors and vitamins	-0.239
DAGPYP_SYN	diacylglycerol pyrophosphate phosphatase (Synechococcus)	Fatty Acid Synthesis	-0.239
ILEDIFF	L-isoleucine transport out via diffusion	Transport	-0.239
LEUDIFF	L-leucine transport out via diffusion	Transport	-0.239
VALDIFF	L-valine transport out via diffusion	Transport	-0.239

Table S11: **Top ten Pearson correlation coefficients between Biomass- Photosystem I fluxes (x) and growth rates (y), Related to Figures 5 and S3.**

Biomass - Photosystem II			
Reaction	Description	Subsystem	PCC
SUCCt2b	Succinate efflux via proton symport	Transport	0.373
EX_succ_E	Succinate exchange	Exchange	0.373
ASPT	L-aspartase	AA Metabolism	0.364
NADPHPQ9tlm	NADPH dehydrogenase (plastoquinone-9 and 4 protons)	Energy Metabolism	0.354
HDH2	Bidirectional Hydrogenase (NADP/NADPH)	Hydrogen Metabolism	0.354
NAT3	sodium transport out via proton antiport	Transport	0.344
GLUSYM	L-Glutamate transport in via sodium symport	Transport	0.343
EX_GLU_E	glutamate exchange	Exchange	0.343
SSALY	succinate-semialdehyde dehydrogenase (NADP)	AA Metabolism and Carbohydrate Metabolism	0.342
2OGDC	2-oxoglutarate decarboxylase	Exchange	0.342
PSIIR	photosystem II reaction	Energy Metabolism	-0.319
O2EXtll	oxygen exchange between thylakoid lumen and cytosol	None	-0.319
H2OEXtll	water exchange between thylakoid lumen and cytosol	None	-0.319
PFK	phosphofructokinase	Carbohydrate Metabolism	-0.239
FRUK	fructokinase	Carbohydrate Metabolism	-0.239
SPS	sucrose-phosphate synthase	Carbohydrate Metabolism	-0.239
SUCPPT	sucrose-phosphate phosphatase	Carbohydrate Metabolism	-0.239
SUCPR	sucrose phosphorylase	Carbohydrate Metabolism	-0.239
GLCP	glycogen phosphorylase	Carbohydrate Metabolism	-0.239
CTPD	CTP deaminase	Nucleotide Metabolism	-0.239

Table S12: **Top ten Pearson correlation coefficients between fluxes for Biomass - Photosystem II (x) and growth rates (y), Related to Figures 5 and S4.**

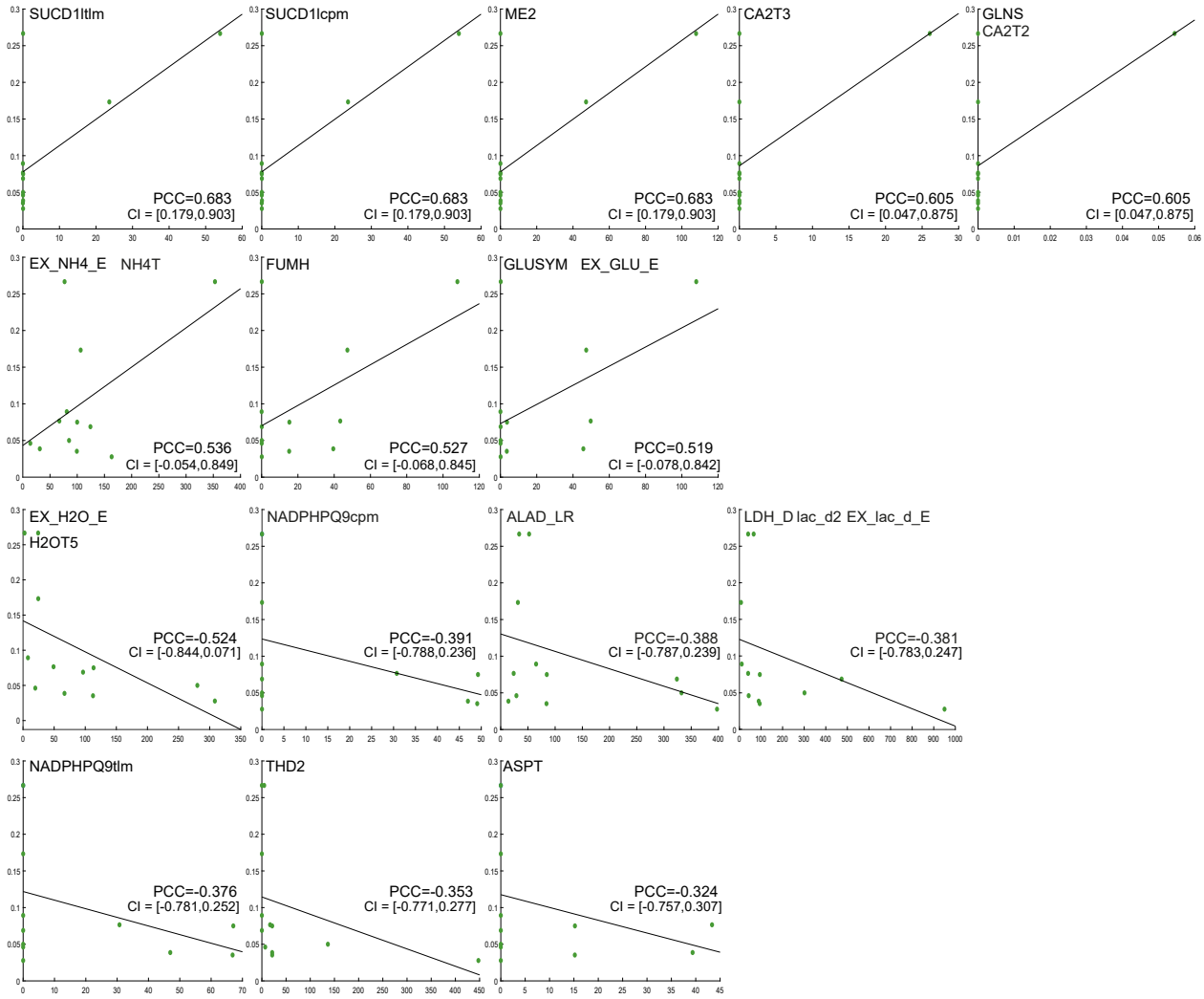


Figure S2: **Top ten Pearson correlation coefficients (PCC) with their respective 95% confidence intervals (CI) between fluxes obtained using Biomass - ATP maintenance as objective functions (x) and growth rates (y), Related to Figure 5.** A list of each reaction and its respective PCC is provided in Table S10.

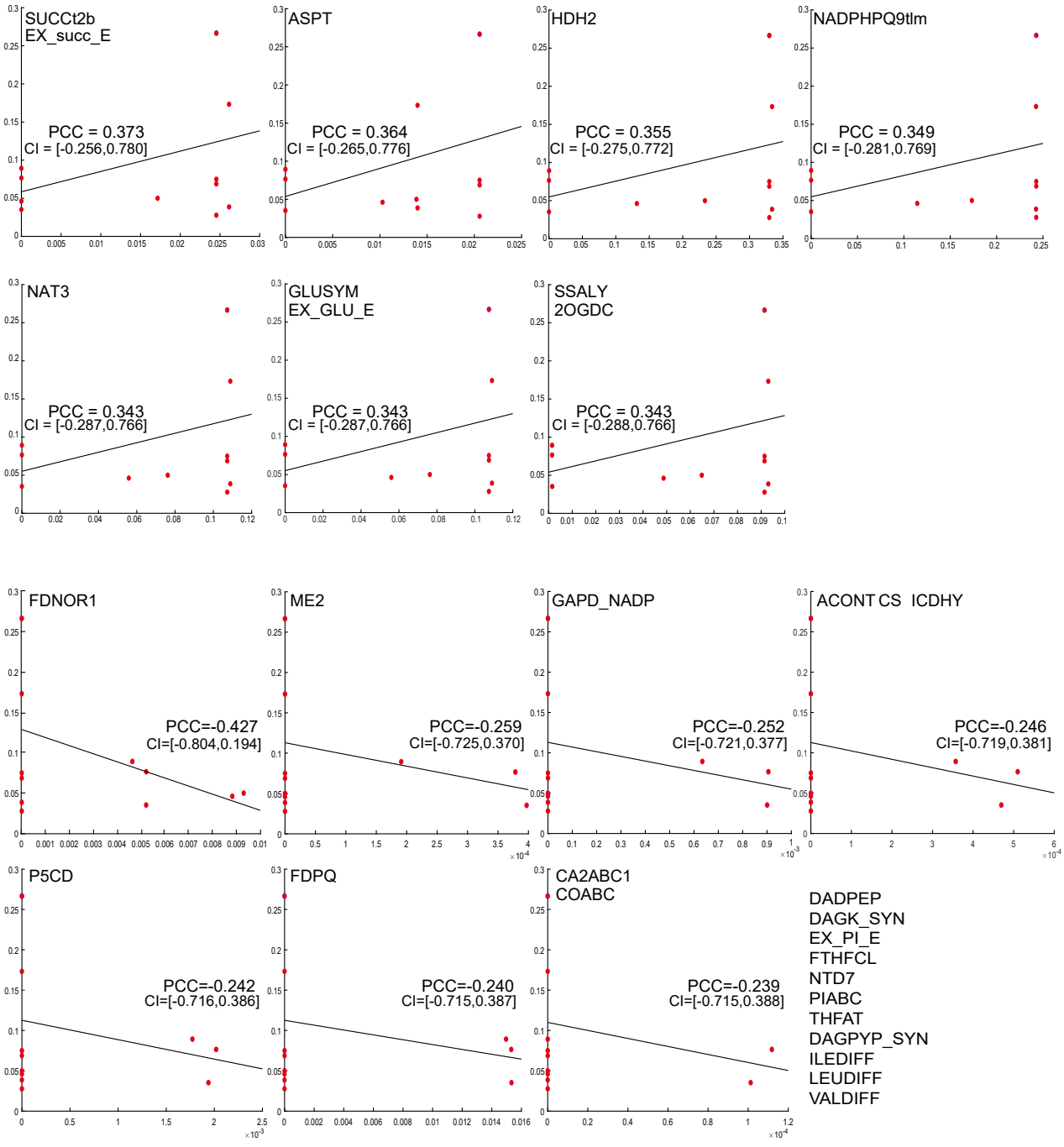


Figure S3: **Top ten Pearson correlation coefficients (PCC) with their respective 95% confidence intervals (CI) between fluxes obtained using Biomass - Photosystem I as objective functions (x) and growth rates (y), Related to Figure 5.** A list of each reaction and its respective PCC is provided in Table S11.

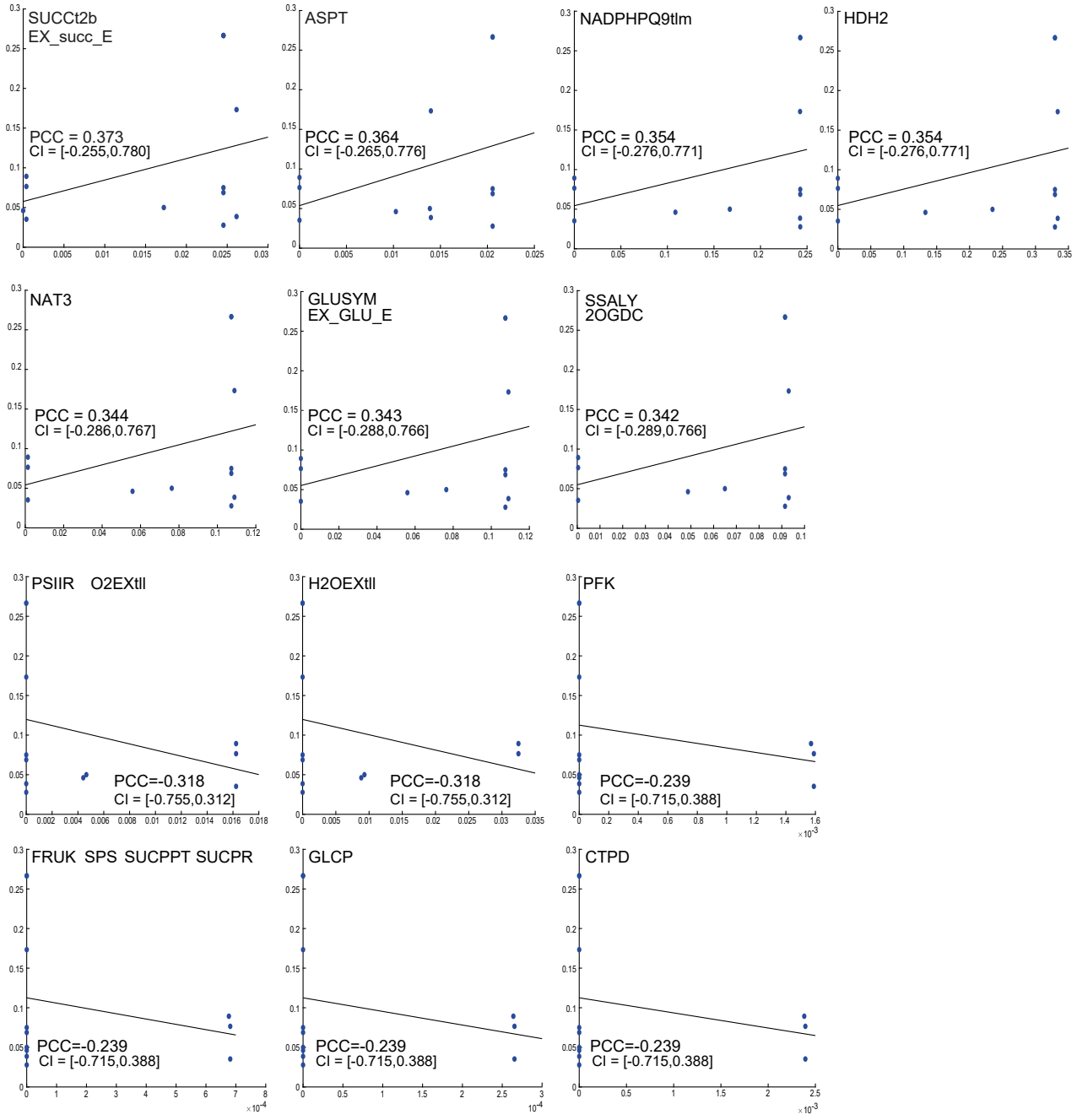


Figure S4: **Top ten Pearson correlation coefficients (PCC) with their respective 95% confidence intervals (CI) between fluxes obtained using Biomass - Photosystem II as objective functions (x) and growth rates (y), Related to Figure 5.** A list of each reaction and its respective PCC is provided in Table S12.

Biomass - ATP maintenance

Gene	COG category	CY category	CY subcategory	PCC
A0639	None	Photosynthesis and respiration	Phycobilisome	0.874
A2457	Post-translational modification, protein turnover, chaperones	Cellular processes	Chaperones	0.828
A1008	None	Photosynthesis and respiration	Photosystem I	0.821
E0022	None	Unknown	NA	0.811
A0561	None	Hypothetical	NA	0.770
A0575	None	Hypothetical	NA	0.767
A2725	None	NA	NA	0.759
A1802	Energy production and conversion/Secondary metabolites biosynthesis, transport and metabolism	Photosynthesis and respiration	CO_2 fixation	0.742
A0881	AA transport and metabolism	AA biosynthesis	Aspartate family	0.734
A0472	None	Hypothetical	NA	0.727
G0060	Inorganic ion transport and metabolism	Transport and binding proteins	NA	-0.835
A2485	General	Hypothetical	NA	-0.831
A0356	General	Translation	Degradation of proteins peptides and glycopeptides	-0.787
A1839	None	Other	Other	-0.781
A1133	Replication, recombination, and repair	Translation	Aminoacyl tRNA synthetases and tRNA modification	-0.766
A1618	AA transport and metabolism	Hypothetical	NA	-0.761
A1703	AA transport and metabolism	Energy Metabolism	Amino acids and amines	-0.759
A2018	General	Hypothetical	NA	-0.758
A1427	Coenzyme transport and metabolism	Biosynthesis of cofactors, prosthetic groups, and carriers	Pantothenate	-0.757
A0483	General	Hypothetical	NA	-0.751

Biomass - Photosystem I

Gene	COG category	CY category	CY subcategory	PCC
A0639	None	Photosynthesis and respiration	Phycobilisome	0.874
A2457	Post-translational modification, protein turnover, chaperones	Cellular processes	Chaperones	0.828
A1008	None	Photosynthesis and respiration	Photosystem I	0.821
E0022	None	Unknown	NA	0.811
A0561	None	Hypothetical	NA	0.770
A0575	None	Hypothetical	NA	0.767
A2725	None	NA	NA	0.759
A1802	Energy production and conversion/Secondary metabolites biosynthesis, transport and metabolism	Photosynthesis and respiration	CO_2 fixation	0.742
A0881	AA transport and metabolism	AA biosynthesis	Aspartate family	0.734
A0472	None	Hypothetical	NA	0.727
G0060	Inorganic ion transport and metabolism	Transport and binding proteins	NA	-0.835
A2485	General	Hypothetical	NA	-0.831
A0356	General	Translation	Degradation of proteins peptides and glycopeptides	-0.787
A1839	None	Other	Other	-0.781
A1133	Replication, recombination, and repair	Translation	Aminoacyl tRNA synthetases and tRNA modification	-0.766
A1618	AA transport and metabolism	Hypothetical	NA	-0.761
A1703	AA transport and metabolism	Energy Metabolism	Amino acids and amines	-0.759
A2018	General	Hypothetical	NA	-0.758
A1427	Coenzyme transport and metabolism	Biosynthesis of cofactors, prosthetic groups, and carriers	Pantothenate	-0.757
A0483	General	Hypothetical	NA	-0.751

Biomass - Photosystem II

Gene	COG category	CY category	CY subcategory	PCC
A0639	None	Photosynthesis and respiration	Phycobilisome	0.874
A2457	Post-translational modification, protein turnover, chaperones	Cellular processes	Chaperones	0.828
A1008	None	Photosynthesis and respiration	Photosystem I	0.821
E0022	None	Unknown	NA	0.811
A0561	None	Hypothetical	NA	0.770
A0575	None	Hypothetical	NA	0.767
A2725	None	NA	NA	0.759
A1802	Energy production and conversion/Secondary metabolites biosynthesis, transport and metabolism	Photosynthesis and respiration	CO_2 fixation	0.742
A0881	AA transport and metabolism	AA biosynthesis	Aspartate family	0.734
A0472	None	Hypothetical	NA	0.727
G0060	Inorganic ion transport and metabolism	Transport and binding proteins	NA	-0.835
A2485	General	Hypothetical	NA	-0.831
A0356	General	Translation	Degradation of proteins peptides and glycopeptides	-0.787
A1839	None	Other	Other	-0.781
A1133	Replication, recombination, and repair	Translation	Aminoacyl tRNA synthetases and tRNA modification	-0.766
A1618	AA transport and metabolism	Hypothetical	NA	-0.761
A1703	AA transport and metabolism	Energy Metabolism	Amino acids and amines	-0.759
A2018	General	Hypothetical	NA	-0.758
A1427	Coenzyme transport and metabolism	Biosynthesis of cofactors, prosthetic groups, and carriers	Pantothenate	-0.757
A0483	General	Hypothetical	NA	-0.751

Table S13: **Top ten Pearson correlation coefficients between transcript-and-flux data (x) and growth rates (y), Related to Figures 5 and S1.** The same genes are already highlighted in Table S9.

References

Aikawa, S., Nishida, A., Ho, S.H., Chang, J.S., Hasunuma, T., Kondo, A., 2014. Glycogen production for biofuels by the euryhaline cyanobacteria *Synechococcus* sp. strain pcc 7002 from an oceanic environment. Biotechnology for biofuels 7, 88.

Angione, C., Conway, M., Lió, P., 2016. Multiplex methods provide effective integration of multi-omic data in genome-scale models. BMC bioinformatics 17, 257.

Angione, C., Lió, P., 2015. Predictive analytics of environmental adaptability in multi-omic network models. Scientific reports 5, 15147.

Bartsevich, V.V., Pakrasi, H., 1995. Molecular identification of an abc transporter complex for manganese: analysis of a cyanobacterial mutant strain impaired in the photosynthetic oxygen evolution process. The EMBO journal 14, 1845–1853.

Boulay, C., Abasova, L., Six, C., Vass, I., Kirilovsky, D., 2008. Occurrence and function of the orange carotenoid protein in photoprotective mechanisms in various cyanobacteria. Biochimica et Biophysica Acta (BBA)-Bioenergetics 1777, 1344–1354.

Checchetto, V., Segalla, A., Allorent, G., La Rocca, N., Leanza, L., Giacometti, G.M., Uozumi, N., Finazzi, G., Bergantino, E., Szabò, I., 2012. Thylakoid potassium channel is required for efficient photosynthesis in cyanobacteria. Proceedings of the National Academy of Sciences 109, 11043–11048.

Clark, R.L., McGinley, L.L., Purdy, H.M., Korosh, T.C., Reed, J.L., Root, T.W., Pfleger, B.F., 2018. Light-optimized growth of cyanobacterial cultures: Growth phases and productivity of biomass and secreted molecules in light-limited batch growth. Metabolic engineering 47, 230–242.

Culley, C., Vijayakumar, S., Zampieri, G., Angione, C., 2020. A mechanism-aware and multiomic machine-learning pipeline characterizes yeast cell growth. Proceedings of the National Academy of Sciences 117, 18869–18879.

Demtröder, L., Narberhaus, F., Masepohl, B., 2019. Coordinated regulation of nitrogen fixation and molybdate transport by molybdenum. Molecular microbiology 111, 17–30.

Ebrahim, A., Brunk, E., Tan, J., O'brien, E.J., Kim, D., Szubin, R., Lerman, J.A., Lechner, A., Sastry, A., Bordbar, A., et al., 2016. Multi-omic data integration enables discovery of hidden biological regularities. Nature communications 7, 1–9.

Flores, E., Frías, J.E., Rubio, L.M., Herrero, A., 2005. Photosynthetic nitrate assimilation in cyanobacteria. Photosynthesis Research 83, 117–133.

Fujisawa, T., Narikawa, R., Maeda, S.i., Watanabe, S., Kanesaki, Y., Kobayashi, K., Nomata, J., Hanaoka, M., Watanabe, M., Ehira, S., et al., 2016. Cyanobase: a large-scale update on its 20th anniversary. Nucleic acids research 45, D551–D554.

Heikal, A., Nakatani, Y., Dunn, E., Weimar, M.R., Day, C.L., Baker, E.N., Lott, J.S., Sazanov, L.A., Cook, G.M., 2014. Structure of the bacterial type ii nadh dehydrogenase: a monotopic membrane protein with an essential role in energy generation. Molecular microbiology 91, 950–964.

Heirendt, L., Arreckx, S., Pfau, T., Mendoza, S.N., Richelle, A., Heinken, A., Haraldsdóttir, H.S., Wachowiak, J., Keating, S.M., Vlasov, V., et al., 2019. Creation and analysis of biochemical constraint-based models using the cobra toolbox v. 3.0. Nature protocols 14, 639–702.

Hendry, J.I., Prasannan, C.B., Joshi, A., Dasgupta, S., Wangikar, P.P., 2016. Metabolic model of *Synechococcus* sp. pcc 7002: Prediction of flux distribution and network modification for enhanced biofuel production. Bioresource Technology 213, 190–197.

Huertas, M.J., López-Maury, L., Giner-Lamia, J., Sánchez-Riego, A.M., Florencio, F.J., 2014. Metals in cyanobacteria: analysis of the copper, nickel, cobalt and arsenic homeostasis mechanisms. Life 4, 865–886.

Jackson, S.A., Eaton-Rye, J.J., Bryant, D.A., Posewitz, M.C., Davies, F.K., 2015. Dynamics of photosynthesis in a glycogen-deficient glgc mutant of *Synechococcus* sp. strain pcc 7002. Appl. Environ. Microbiol. 81, 6210–6222.

Kato, A., Takatani, N., Ikeda, K., Maeda, S.i., Omata, T., 2017. Removal of the product from the culture medium strongly enhances free fatty acid production by genetically engineered *Synechococcus elongatus*. Biotechnology for biofuels 10, 141.

Kim, J., Fabris, M., Baart, G., Kim, M.K., Goossens, A., Vyverman, W., Falkowski, P.G., Lun, D.S., 2016. Flux balance analysis of primary metabolism in the diatom phaeodactylum tricornutum. The Plant Journal 85, 161–176.

Lê, S., Josse, J., Husson, F., et al., 2008. Factominer: an r package for multivariate analysis. Journal of statistical software 25, 1–18.

Lea-Smith, D.J., Bombelli, P., Vasudevan, R., Howe, C.J., 2016. Photosynthetic, respiratory and extracellular electron transport pathways in cyanobacteria. Biochimica et Biophysica Acta (BBA)-Bioenergetics 1857, 247–255.

Liu, X., Yang, M., Wang, Y., Chen, Z., Zhang, J., Lin, X., Ge, F., Zhao, J., 2018. Effects of psii manganese-stabilizing protein succinylation on photosynthesis in the model cyanobacterium *Synechococcus* sp. pcc 7002. Plant and Cell Physiology 59, 1466–1482.

Ludwig, M., Bryant, D.A., 2011. Transcription profiling of the model cyanobacterium *Synechococcus* sp. strain pcc 7002 by next-gen (solid™) sequencing of cdna. Frontiers in microbiology 2, 41.

Ludwig, M., Bryant, D.A., 2012a. Acclimation of the global transcriptome of the cyanobacterium *Synechococcus* sp. strain pcc 7002 to nutrient limitations and different nitrogen sources. Frontiers in microbiology 3, 145.

Ludwig, M., Bryant, D.A., 2012b. *Synechococcus* sp. strain pcc 7002 transcriptome: acclimation to temperature, salinity, oxidative stress, and mixotrophic growth conditions. Frontiers in microbiology 3, 354.

McConnell, M.D., Cowgill, J.B., Baker, P.L., Rappaport, F., Redding, K.E., 2011. Double reduction of plastoquinone to plastoquinol in photosystem 1. Biochemistry 50, 11034–11046.

McNeely, K., Xu, Y., Bennette, N., Bryant, D.A., Dismukes, G.C., 2010. Redirecting reductant flux into hydrogen production via metabolic engineering of fermentative carbon metabolism in a cyanobacterium. Applied and environmental microbiology 76, 5032–5038.

Myers, J.A., Curtis, B.S., Curtis, W.R., 2013. Improving accuracy of cell and chromophore concentration measurements using optical density. BMC biophysics 6, 4.

Noreña-Caro, D., Benton, M.G., 2018. Cyanobacteria as photoautotrophic biofactories of high-value chemicals. Journal of CO2 Utilization 28, 335–366.

Palenik, B., Brahamsha, B., Larimer, F., Land, M., Hauser, L., Chain, P., Lamerdin, J., Regala, W., Allen, E., McCarren, J., et al., 2003. The genome of a motile marine *Synechococcus*. Nature 424, 1037–1042.

Peltier, G., Aro, E.M., Shikanai, T., 2016. Ndh-1 and ndh-2 plastoquinone reductases in oxygenic photosynthesis. Annual review of plant biology 67, 55–80.

Qian, X., Kumaraswamy, G.K., Zhang, S., Gates, C., Ananyev, G.M., Bryant, D.A., Dismukes, G.C., 2016. Inactivation of nitrate reductase alters metabolic branching of carbohydrate fermentation in the cyanobacterium *Synechococcus* sp. strain pcc 7002. Biotechnology and bioengineering 113, 979–988.

Qiao, C., Duan, Y., Zhang, M., Hagemann, M., Luo, Q., Lu, X., 2018. Effects of reduced and enhanced glycogen pools on salt-induced sucrose production in a sucrose-secreting strain of *Synechococcus elongatus* pcc 7942. Appl. Environ. Microbiol. 84, e02023–17.

Saha, R., Liu, D., Hoynes-O'Connor, A., Liberton, M., Yu, J., Bhattacharyya-Pakrasi, M., Balassy, A., Zhang, F., Moon, T.S., Maranas, C.D., et al., 2016. Diurnal regulation of cellu-

lar processes in the cyanobacterium *Synechocystis* sp. strain pcc 6803: Insights from transcriptomic, fluxomic, and physiological analyses. MBio 7, e00464–16.

Sarkar, D., Mueller, T.J., Liu, D., Pakrasi, H.B., Maranas, C.D., 2019. A diurnal flux balance model of *Synechocystis* sp. pcc 6803 metabolism. PLOS Computational Biology 15, e1006692.

Shcolnick, S., Keren, N., 2006. Metal homeostasis in cyanobacteria and chloroplasts. balancing benefits and risks to the photosynthetic apparatus. Plant physiology 141, 805–810.

Shvarev, D., Maldener, I., 2018. Atp-binding cassette transporters of the multicellular cyanobacterium anabaena sp. pcc 7120: A wide variety for a complex lifestyle. FEMS microbiology letters 365, fny012.

Ueda, K., Nakajima, T., Yoshikawa, K., Toya, Y., Matsuda, F., Shimizu, H., 2018. Metabolic flux of the oxidative pentose phosphate pathway under low light conditions in *Synechocystis* sp. pcc 6803. Journal of bioscience and bioengineering 126, 38–43.

Vu, T.T., Stolyar, S.M., Pinchuk, G.E., Hill, E.A., Kucek, L.A., Brown, R.N., Lipton, M.S., Osterman, A., Fredrickson, J.K., Konopka, A.E., et al., 2012. Genome-scale modeling of light-driven reductant partitioning and carbon fluxes in diazotrophic unicellular cyanobacterium *Cyanothece* sp. atcc 51142. PLoS computational biology 8, e1002460.

Watanabe, M., Semchonok, D.A., Webber-Birungi, M.T., Ehira, S., Kondo, K., Narikawa, R., Ohmori, M., Boekema, E.J., Ikeuchi, M., 2014. Attachment of phycobilisomes in an antenna–photosystem i supercomplex of cyanobacteria. Proceedings of the National Academy of Sciences 111, 2512–2517.

Woo, J.E., Jang, Y.S., 2019. Metabolic engineering of microorganisms for the production of ethanol and butanol from oxides of carbon. Applied Microbiology and Biotechnology 103, 8283–8292.

Yang, J.H., Wright, S.N., Hamblin, M., McCloskey, D., Alcantar, M.A., Schrübbers, L., Lopatkin, A.J., Satish, S., Nili, A., Palsson, B.O., et al., 2019. A white-box machine learning approach for revealing antibiotic mechanisms of action. Cell 177, 1649–1661.

Yang, Y., Feng, J., Li, T., Ge, F., Zhao, J., 2015. Cyanomics: an integrated database of omics for the model cyanobacterium *Synechococcus* sp. pcc 7002. Database 2015, bau127.

Zampieri, G., Vijayakumar, S., Yaneske, E., Angione, C., 2019. Machine and deep learning meet genome-scale metabolic modeling. PLoS computational biology 15.

Zhang, J., Petersen, S.D., Radivojevic, T., Ramirez, A., Pérez-Manríquez, A., Abeliuk, E., Sánchez, B.J., Costello, Z., Chen, Y., Fero, M.J., et al., 2020. Combining mechanistic and machine learning models for predictive engineering and optimization of tryptophan metabolism. Nature Communications 11, 1–13.