

High-Throughput Metabolic Profiling for Model Refinements of Microalgae

Amnah Alzahmi^{1,2}, Sarah Daakour¹, Diana Charles El Assal³, Bushra S. Dohai³, Amphun Chaiboonchoe³, Weiqi Fu^{3,4}, David R. Nelson¹, Alexandra Mystikou¹, Kourosh Salehi-Ashtiani^{1,3}

¹ Center for Genomics and Systems Biology (CGSB), New York University Abu Dhabi Research Institute ² Department of Biology, United Arab Emirates University ³ Laboratory of Algal, Systems, and Synthetic Biology (LASSB), Division of Science and Math, New York University Abu Dhabi ⁴ Department of Marine Science, Ocean College, Zhejiang University

Corresponding Author

Kourosh Salehi-Ashtiani

ksa3@nyu.edu

Citation

Alzahmi, A., Daakour, S., El Assal, D.C., Dohai, B.S., Chaiboonchoe, A., Fu, W., Nelson, D.R., Mystikou, A., Salehi-Ashtiani, K. High-Throughput Metabolic Profiling for Model Refinements of Microalgae. *J. Vis. Exp.* (178), e61913, doi:10.3791/61913 (2021).

Date Published

December 4, 2021

DOI

10.3791/61913

URL

joVE.com/video/61913

Abstract

Metabolic models are reconstructed based on an organism's available genome annotation and provide predictive tools to study metabolic processes at a systems-level. Genome-scale metabolic models may include gaps as well as reactions that are unverified experimentally. Reconstructed models of newly isolated microalgal species will result in weaknesses due to these gaps, as there is usually sparse biochemical evidence available for the metabolism of such isolates. The phenotype microarray (PM) technology is an effective, high-throughput method that functionally determines cellular metabolic activities in response to a wide array of entry metabolites. Combining the high throughput phenotypic assays with metabolic modeling can allow existing metabolic network models to be rapidly reconstructed or optimized by providing biochemical evidence to support and expand genomic evidence. This work will show the use of PM assays for the study of microalgae by using the green microalgal model species *Chlamydomonas reinhardtii* as an example. Experimental evidence for over 254 reactions obtained by PM was used in this study to expand and refine a genome-scale *C. reinhardtii* metabolic network model, *i*RC1080, by approximately 25 percent. The protocol created here can be used as a basis for functionally profiling the metabolism of other microalgae, including known microalgae mutants and new isolates.

Introduction

Optimizing algal metabolism for enhanced and stable production of targeted metabolites requires the development of complex metabolic engineering strategies through systems-level analyses of metabolic networks. Metabolic

network models can guide the rational designs for the rapid development of optimization strategies^{1,2,3,4}. Although approximately 160 microalgal species have been sequenced⁵, there are, to our knowledge, only 44 algal

metabolic models available^{4,6,7}. Due to the difficulty in obtaining high-throughput metabolic phenotypic data for experimental validation of genomic information, the reconstruction of high-quality network models lags behind the rapid development of algal genome sequencing.

C. reinhardtii is an attractive model system for algal-based studies. This species can grow photoautotrophically or heterotrophically and has been widely used as a model organism in basic and applied research. Its genome sequence was published in 2007⁸, with genome-scale metabolic models subsequently reconstructed for the species^{9,10,11}. The genome-scale model for *C. reinhardtii* (iRC1080) was reconstructed by Chang *et al.*¹⁰ based on genomic and literature evidence (entailing ~250 sources). It has 1,706 metabolites with 2,190 reactions¹⁰; however, the completeness of the model could not be verified beyond the available published experimental evidence at the time.

The phenotype microarrays (PMs) technology is a high-throughput platform that can provide metabolic profiling information for heterotrophic microorganisms as well as tissue-culture cells. In particular, it can be used to address the phenotype-to-genotype knowledge gap in microalgae, as first reported for *Chlamydomonas reinhardtii*¹² and subsequently for a species of Chloroidium¹³ and Chlorella¹⁴. By studying cell responses to thousands of metabolites, signaling molecules, osmolytes, and effector molecules, the PM assays can provide functional metabolic profiling and offer insights into the function, metabolism, and environmental sensitivity^{15,16,17}. Specifically, PM assays detect cells metabolite utilization in 96-well microplates with different nutrients, metabolites, or osmolytes contained in each well. Moreover, it is also possible to assay bioactive molecules, such as antibiotics and hormones.

As determined by the intensity of color production by the NADH reduction of a tetrazolium-based redox dye, the metabolic utilization of substrates is evaluated in terms of cell respiration^{15,16,17}. The experiments in 96-well microplates can be monitored and determined automatically over time with the phenotype microarray instrument (PMI) platform. Twenty 96-well microplates are designed to represent the common set metabolites to study cellular phenotypes to utilize carbon, nitrogen, sulfur, and phosphorus sources, along with different osmotic/ion and pH effects. The PM technology has been successfully used for updating and upgrading a number of existing genome-scale metabolic models for microorganisms^{15,16,17,18}.

The protocol and data shown here are based on previously published work by Chaiboonchoe *et al.*¹² The presented work details the use of the PM assay method to characterize the metabolic phenotypes of microalgae and to expand an existing algal metabolic model of *C. reinhardtii* as well as to guide the reconstruction of new metabolic models.

Protocol

1. Phenotype Microarray Experiments

1. Obtain *C. reinhardtii* strain CC-503 from the Chlamydomonas Resource Center at the University of Minnesota, USA (<https://www.chlamycollection.org>).
2. Grow the cells in fresh Tris-Acetate-Phosphate (TAP) media¹⁹ with final concentrations of 400 µg/mL timentin, 50 µg/mL ampicillin, and 100 µg/mL kanamycin (to inhibit bacterial growth) under 400 micromol photons/m²s, at 25 °C, for two days to mid-log phase.
3. Spin down the culture at 2,000 x g for 10 min at 22 °C and discard the supernatant without disturbing the pellet.

4. Prepare fresh TAP media containing 0.1% tetrazolium violet dye "D".

NOTE: Modify TAP media in this step to exclude some nutrients depending on the metabolite category tested in each plate (e.g., exclude ammonium chloride for nitrogen source plates).

5. Resuspend the pellet in fresh TAP media prepared (from step 1.2) to a final concentration of 1×10^6 cells/mL.
6. Use chemical compound array assay plates (carbon sources, nitrogen sources, phosphorus and sulfur sources plates, and the peptide nitrogen sources).
7. Inoculate a 100 μ L aliquot of cell-containing media into each well of the assay plates.

NOTE: Make sure to duplicate the assays.

8. Streak cells on yeast extract/peptone plates and perform gram staining, as in Smith *et al.*²⁰ before and after the assay to monitor bacterial contamination.
9. Insert the chemical compound array assay plates into the microplate reader system.
10. Incubate all the plates at 30 °C for up to 7 days and program the microplate reader system to read the dye color change every 15 min.

NOTE: As most microplate readers do not provide a source of continuous light during incubation, the algae should be able to carry out heterotrophic respiration.

2. Data Analysis

1. Export the raw kinetic data from the microplate reader as CSV files, which will subsequently be used as input to the Omnilog Phenotype Microarray (OPM) package in R. Add the biological information as metadata (e.g., strain designation, growth media, temperature, etc.).

1. Using the PM Kinetic data converter software; load the D5E data files, and convert them to OKA files using the following command lines in the PM kinetic analysis software:

Load | Import (locate the folder of the OKA files) | Populate Filters | Import | Add All Plates | Close. Export | choose read data (Kinetic), choose format (CSV) (Tabulate Header), and choose plates (every plate (individual Files)) | Export data | Save.

2. To carry out the Phenotype Microarray (PM) data analysis, use the OPM software package^{21,22} that runs within the R software environment. The package, tutorial, and reference documentation are available at: <http://www.goeker.org/opm/>. In RStudio, a graphical user interface for R, install the opm package and its dependencies using the following commands:

```
source (http://www.goeker.org/opm/
install_opm.R)
library (opm)
```

3. Navigate to the directory that contains the CSV files of the kinetic data and import the data using the `read_opm` function:

```
x <- read_opm(".", convert="grp", include=list
("csv:"))
```

4. Aggregate and discretize the kinetic data using curve-parameter estimation.

```
For (i in 1:length(x)) {
x[[i]] <-l do _aggre(x[[i]], boot = 0 L,
cores = 1 L, method ="splines", options =
set_spline_options("smooth.spline"))
x[[i]] <- do_disc(x[[i]], cutoff = FALSE)
```

```

}
#Collection of the metadata
metadata <- collect_template(".")
metadata$Strain <- c("BLANK","CC- 503")
for (l in 1 :length(x)) {x[[l]] <-
include_metadata(x[[l]], md = metadata, replace
= TRUE)}

```

5. Use the function `xy_plot` to map the respiration (or growth) measurements (y-axis) as a function of time (x-axis) for the assayed 96-well plates.

```

print (xy_plot(x[[ 1 ]], include ="Strain",
theor.max = FALSE))

```

6. Visualize the data as a heat map using the function `level_plot` to allow for a quick comparative overview of the kinetic data.

```

level_plot(x, main = list(), colors =
opm_opt("color.borders"), panel.headers =
metadata$Strain, cex = NULL, strip.fmt = list(),
striptext.fmt = list(), legend.sep = " ", space
="Lab", bias = 0.7 , num.colors = 200 L)

```

7. Extract important biological information, the *curve parameters*, from the raw kinetic curves and include the lag phase (λ), the growth rate (μ), the maximum cell respiration (A), and the area under the curve (AUC)²¹. To identify positive metabolites, use the A values of the negative control, which represents the abiotic reactivity of the dye with the medium, in addition to the blank of each microwell plate as background subtraction values. The *extract* function is used to obtain the A parameter.

```

opm_opt("curve.param")
param <- extract (x, as.labels = list("Strain"))

```

3. Identification of Reactions and Genes Associated with New Metabolites

1. Search KEGG (Kyoto Encyclopedia of Genes and Genomes) (<http://www.genome.jp/kegg/>) and MetaCyc (<http://metacyc.org/>) to identify Enzyme Commission numbers (ECs) for reactions using metabolites found from chemical compound arrays^{23,24}.
2. Use the identified EC numbers as a search basis in multiple available algal annotation resources such as Joint Genome Institute (JGI), Phytozome (<http://www.phytozome.net/>), and peer-reviewed publications^{23,25,26,27}.
3. If a query returns no genetic evidence for a given EC number, identify the relevant associated proteins in other organisms, starting with species closest to the *C. reinhardtii*, then carry out a profile-based search using the NCBI PSI-BLAST server with default settings and use non-redundant proteins (nr) in *C. reinhardtii* (taxid:3055) to identify candidate genes associated to the reaction¹².
4. Manually curate PSI-BLAST hits with E values of < 0.05 for relevance to the searched EC number through querying those BLAST hits through EMBL-EBI Pfam (<http://pfam.xfam.org/search>), or InterPro (<http://www.ebi.ac.uk/interpro/>) protein domain prediction servers. Note that the latter two scans are critical steps to ensure the identification of the correct enzymatic activity for the protein.

4. Model Refinement and Evaluation

1. Use the latest COBRA Toolbox v.3.0²⁸ in MATLAB^{29,30} platform to carry out the following steps for model refinement. The COBRA Toolbox can be installed

by following the steps in: <https://opencobra.github.io/cobratoolbox/stable/installation.html>. Alternatively, note that the COBRA Toolbox is also implemented across other open-source programming languages, such as Python (COBRAPy³¹) and is available at: <https://opencobra.github.io/cobrapy/>.

1. After installing the COBRA Toolbox v.3.0, open MATLAB and execute the following command to initialize the toolbox:

```
initCobraToolbox;
```

2. Add the identified reactions with their associated genes to the metabolic model, such as *i*RC1080, using the COBRA Toolbox functions *addReaction* and *changeGeneAssociation*. Navigate to the directory that contains the *i*RC1080 model, downloaded from <http://bigg.ucsd.edu/models/iRC1080> and execute the following commands to load the model, rename it, and add a new reaction and its associated gene.

```
Load('iRC 1080 .mat')  
modelNew = iRC 1080;  
modelNew = addReaction(modelNew, 'D-ALA 2',  
...  
{'d-ala[c]', 'atp[c]', ...  
'D-aladata[c]', 'adp[c]', 'pi[c]', ...  
'h[c]'},[- 2 - 1 1 1 1 1 ],false);  
modelNEW =  
changeGeneAssociation(modelNew, ...  
'D-ALA 2','au.g 14655 _t 1' );
```

3. In some cases when the metabolite is not produced intracellularly but is taken up from the medium, add transport reactions for the new metabolites to the model. These transport reactions represent passive diffusion of a metabolite from the

extracellular medium to the cytosol. In addition, add a corresponding artificial exchange reaction using the *addExchangeRxn* function to input or output the metabolite into the extracellular medium.

```
modelNew = addReaction(modelNew, 'CYCPT',  
...  
{'cycp[e]','cycp[c]'},[- 1 1 ],true)  
modelNew = addExchangeRxn(modelNew,  
'cycp[e]', - 1000 ,1000 );
```

4. Test the behavior of the new resultant model, e.g., *i*BD1106, by carrying out flux balance analysis (FBA) using the function *optimizeCbModel* under light and dark conditions for the maximization of biomass as the objective function. For light growth, set the lower and upper bounds of the PRISM solar litho' light reactions to 646.07 (maximum rate). For dark growth, set the bounds of all PRISM light reactions to zero. Use the Biomass function defined previously¹⁰ for growth under dark and light conditions. The FBA solution will output two vectors corresponding to reaction fluxes (solution.v) and reduced cost (solution.w), as well as one vector corresponding to metabolites' shadow prices (solution.y).

%Simulate growth under light condition:

```
modelNew = changeRxnBounds(modelNew,{ ...  
% 'PRISM_solar_litho', ...  
'PRISM_solar_exo', ...  
'PRISM_incandescent_ 60 W', ...  
'PRISM_fluorescent_cool_ 215 W', ...  
'PRISM_metal_halide', ...  
'PRISM_high_pressure_sodium', ...  
'PRISM_growth_room', ...  
'PRISM_white_LED', ...
```

```
'PRISM_red_LED_array_ 653 nm' , ...
'PRISM_red_LED_ 674 nm' , ...
'PRISM_fluorescent_warm_ 18 W' , ...
'PRISM_design_growth' , ...
}, 0, 'b' );
modelNew = changeObjective(modelNew,
'BIOMASS_Chlamy_mixo');
FBAsolutionNew =
optimizeCbModel(modelNew, 'max');
```

- Repeat step 4.1.4 for *i*RC1080 to compare FBA solutions obtained for *i*BD1106 with those obtained for *i*RC1080.
- There is a range of COBRA methods available that can be used to compare models (e.g., flux variability analysis, gene deletion studies, robustness analyses, flux split predictions, FBA, sampling, etc.). Detailed tutorials can be found at <https://opencobra.github.io/cobratoolbox/stable/tutorials/index.html>. Here, an example is provided where the *i*RC1080 model is compared with its refined version, *i*BD1106, by obtaining the shadow prices (sensitivity of the biomass objective function to changes in system variable) of the metabolites accounted for in each model.

Obtain the shadow prices for the metabolites:

```
shadowPrices = table(modelNew.mets, ...
modelNew.metNames, FBAsolutionNew.y);
```

Representative Results

Phenotype Microarray screening of model alga *Chlamydomonas reinhardtii*

The PM assays test the ability of the alga to utilize various sources of carbon, nitrogen, sulfur, and phosphorus in a minimal medium. In this methods description, we

demonstrated how PM assays were used to identify carbon and nitrogen metabolism. Carbon and nitrogen utilization kinetics were measured with a microplate reader. Data were analyzed using PMI software. The summary kinetics of selected PM assay plates (PM01 and PM03) are shown in **Figure 1**. The "xy plots" display the respiration measurements over time plotted for the 96-well plates' assays, where the y-axis and x-axis represent the values of raw measurements and time, respectively. The data was converted to a heat-map pattern to comparatively analyze the assembly of the kinetic data.

The pipeline of refining genome-scale metabolic network using PM data (**Figure 2**) illustrates the integration of the high-throughput PM assays with experimental evidence provided by genomic searches can expand a metabolic network model.

To determine the reproducibility of the PM data obtained from PM01 - 04 and PM10 plates, a linear regression was analyzed to plot the data from two independent replicate experiments against each another (**Figure 3**). **Figure 3** shows that the majority of the data were almost similar as they fall on the 45° line, with only a few outliers being present, and their coefficient of determination R^2 was 0.9. The consistency and reproducibility of the experiments for the alga are verified by this plot.

Identification of new metabolites

The PM assay identified 662 metabolites in seven plates; PM01-PM04 and PM06-PM08, while Gas Chromatography Time-Of-Flight (GC-TOF) had identified 77 metabolites³² (**Figure 4**). When comparing these two sets with the 1068 metabolites accounted in the *i*RC1080, only six metabolites overlapped between the three sets, and 149 overlapped between the *i*RC1080 and the PM. This result

demonstrates that the metabolic profiling platform can be a significant source of new metabolic information.

Acetic acid was the only carbon source detected in plate PM01 as a supporting carbon after subtracting the background signal. This finding is consistent with the literature³³ and shows the specificity of the PM assays. The PM assays revealed new sulfur, phosphorus, and nitrogen sources that *C. reinhardtii* can utilize for growth. The sulfur metabolites were sulfate, thiosulfate, tetrathionate, and DL-Lipoamide. The phosphorus sources were thiophosphate, dithiophosphate, D-3-phospho-glyceric acid, and cysteamine-S-phosphate. The nitrogen source metabolites were L-amino and D-amino acids, including less common amino acids; L-homoserine, L-pyroglutamic, methylamine, ethylamine, ethanolamine, and D,L- α -amino-butyric, and 108 Di-peptides and five Tri-peptides (**Table 1**). All the 128 newly identified metabolites were searched in KEGG and MetaCyc for their associated reactions, EC numbers, and pathways.

The new 128 metabolites were associated with 49 unique EC numbers. Of these, 15 ECs were linked to their genomic evidence using five sources including; Phytozome Version 10.0.2³⁴ JGI Version 4³⁵, AUGUSTUS 5.0, and 5.2¹⁰ annotations from Manichaikul *et al.*³⁶ and KEGG¹³. Metabolites without genomic evidence were entered into the Universal Protein Resource website (UniProt, <http://www.uniprot.org/>)^{37,38} where their related sequences were found in other organisms. Homologous sequences in *C. reinhardtii* were identified by running Position-Specific Iterated BLAST (PSI-BLAST, <https://blast.ncbi.nlm.nih.gov/Blast.cgi>) from the NCBI website considering only sequences that produced significant alignments (E-value <0.005).

Model refinement

Reactions associated with the new 128 metabolites, along with their encoded genes, were added to the *i*RC1080 model, expanding the network. The resulting model *i*BD1106, accounts for 2,444 reactions, 1,959 metabolites, and 1,106 genes (**Table 2**). The new 254 added reactions were 20 amino acid oxidation reactions, 108 di-peptide hydrolysis reactions, five tripeptide hydrolysis reactions, and 120 transport reactions, encoded by four genes (Cre02.g096350.t1.3, au.g14655_t1, e_gwW.1.243.1, Cre12.g486350.t1.3).

A total of 113 added new reactions account for the hydrolysis of di-peptides and tri-peptides. The hydrolysis of di-peptides and tri-peptides are associated with two genes, one for di-peptides (Cre02.g078650.t1.3) and one for tri-peptides (Cre16.g675350.t1.3).

Concerning sources of phosphorus, a reaction for hydrolysis of cysteamine-S-phosphate into cysteamine and phosphate was added associated with the gene JLM_162926.

The WoLF PSORT tool³⁹ (http://www.genscript.com/psort/wolf_psort.html) and results reported by Ghamsari *et al.*³⁵ were applied to obtain the specification of the cellular compartments where the new reactions take place. By analyzing protein sequences associated with the new reactions, WoLF PSORT predicted cytosol as the cellular compartment for the reactions.

A generated metabolic model may contain gaps when the biochemical information is incomplete. In such cases, *gapFind*, A COBRA command, is used. It lists root gaps and allows the identification of new gaps introduction in the new model. The metabolites that cannot be produced in a metabolic model are referred to as root gaps^{40,41}. Analyzing the root gap indicated that both *i*RC1080 and *i*BD1106

models contain the same 91 gaps. This shows that adding the new metabolites and their associated reactions did not introduce any additional root gaps. It should be noted that the phenotyping method used in this protocol does not close root gaps, because the original root gap metabolites lack transport or production mechanisms, which were not addressed in the phenotyping assays. Flux balance analysis was carried out to test the metabolic behavior of *i*BD1106 under light and dark conditions; (no acetate) and (with acetate), respectively. The algorithm maximizes the biomass precursor reactions for an objective function (biomass growth). To evaluate the involvement of each metabolite to the set objective function, "shadow prices" for all metabolites were calculated. The change in the objective function concerning flux changes of

the metabolite defines the shadow price of a metabolite^{30,42}. The indication of whether a metabolite is in "excess" or is "limiting" the objective function can be determined by shadow price analysis, e.g., biomass production. Negative or positive shadow price values reveal metabolites that, upon addition, will decrease or increase the objective function. Zero values of shadow prices reveal metabolites that will not affect the objective function. The comparison of shadow prices between *i*BD1106 and *i*RC1080 in **Figure 5** shows that, for most metabolites, a significant change is not observed; though, differences are found in 105 and 70 cases under light and dark growth conditions, respectively. **Table 4** includes examples of such metabolites.

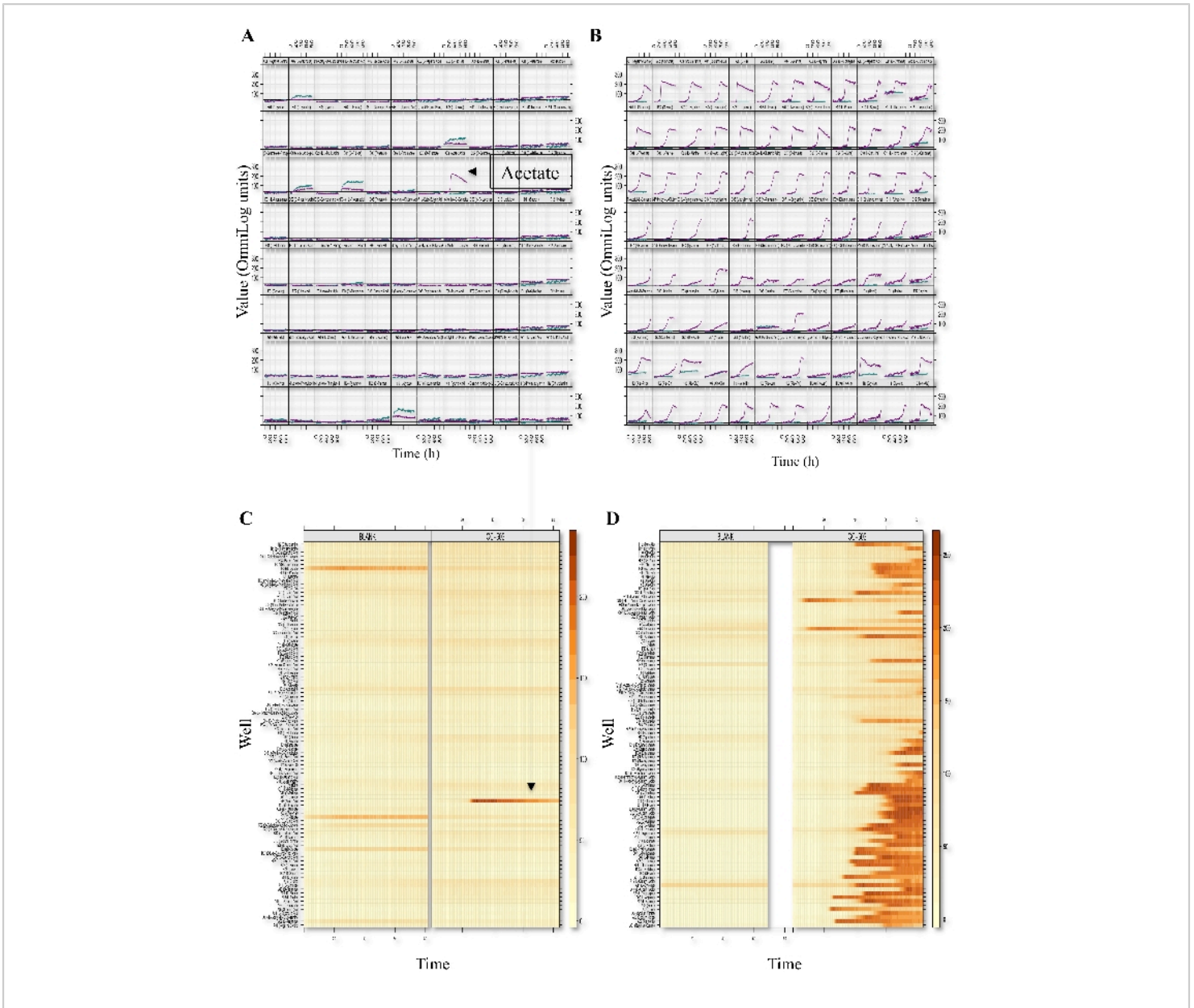


Figure 1: Phenotypic microarray profiling of *C. reinhardtii*. Respiration XY-plots and level plots of the PM01 (Carbon sources; A, C) and PM03 (Nitrogen sources; B, D) assay plates are shown. The figure is an 8x12 array where each cell represents a well-plate and, thus, a given metabolite or growth environment. Within each cell or well representation, curves represent dye conversion by reduction (y-axis) as a function of time (x-axis). PM respiration curves from the CC-503 and blank wells are shown in each cell and are indicated by color (teal color represents blank wells and purple color represents CC-503). The level-plot represents each respiration curve as a thin horizontal line changing color (or remaining unchanged) over time. Heatmap color changes are from light yellow (little to no respiration has taken place) to dark orange or brown

(significant respiration has taken place). Metabolites utilized by *C. reinhardtii* (CC-503) and the blank plates are shown. This figure is from a previously published work by Chaiboonchoe *et al.*¹² [Please click here to view a larger version of this figure.](#)

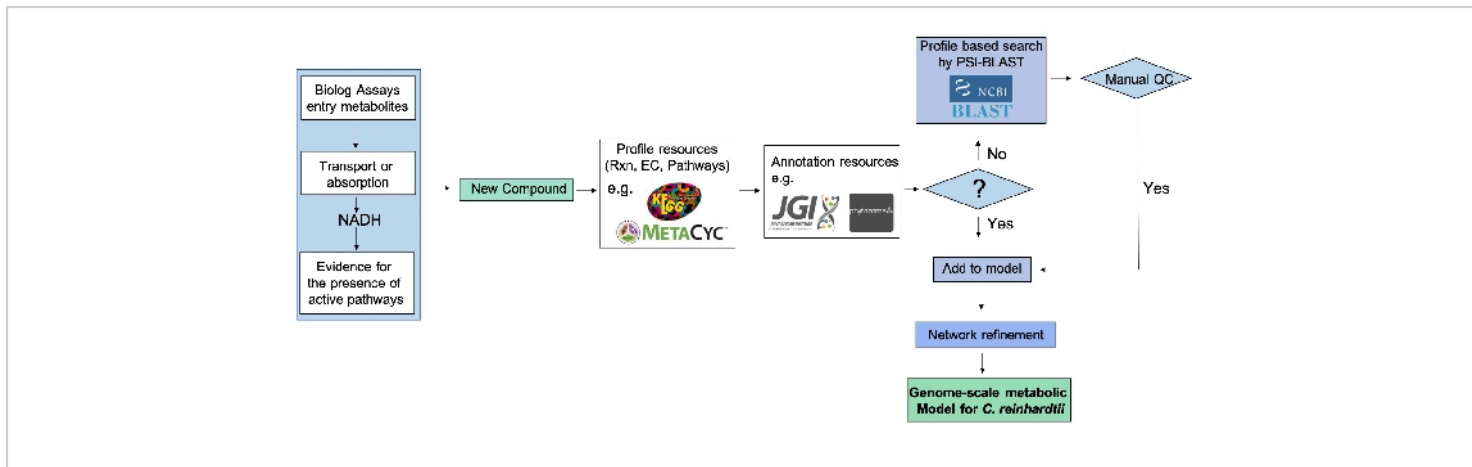


Figure 2: Genome-scale metabolic network refinement pipeline using PM data. After a new compound tests positive in a PM assay, its Enzyme Commission number (EC), reaction, and pathway are identified from available databases, e.g., KEGG and MetaCyc. Genomic evidence is then extracted from genomic and annotation resources when available and constitutes a link between genotype and phenotype. When direct genomic evidence is unavailable, the protein sequence is identified from the EC numbers, and genetic evidence is identified via PSI-Blast. The reconstructed metabolic network is then refined based on newly identified compounds, but only after a quality control step that entails querying the protein domains using relevant databases. This figure has been modified from previously published work by Chaiboonchoe *et al.*¹² [Please click here to view a larger version of this figure.](#)

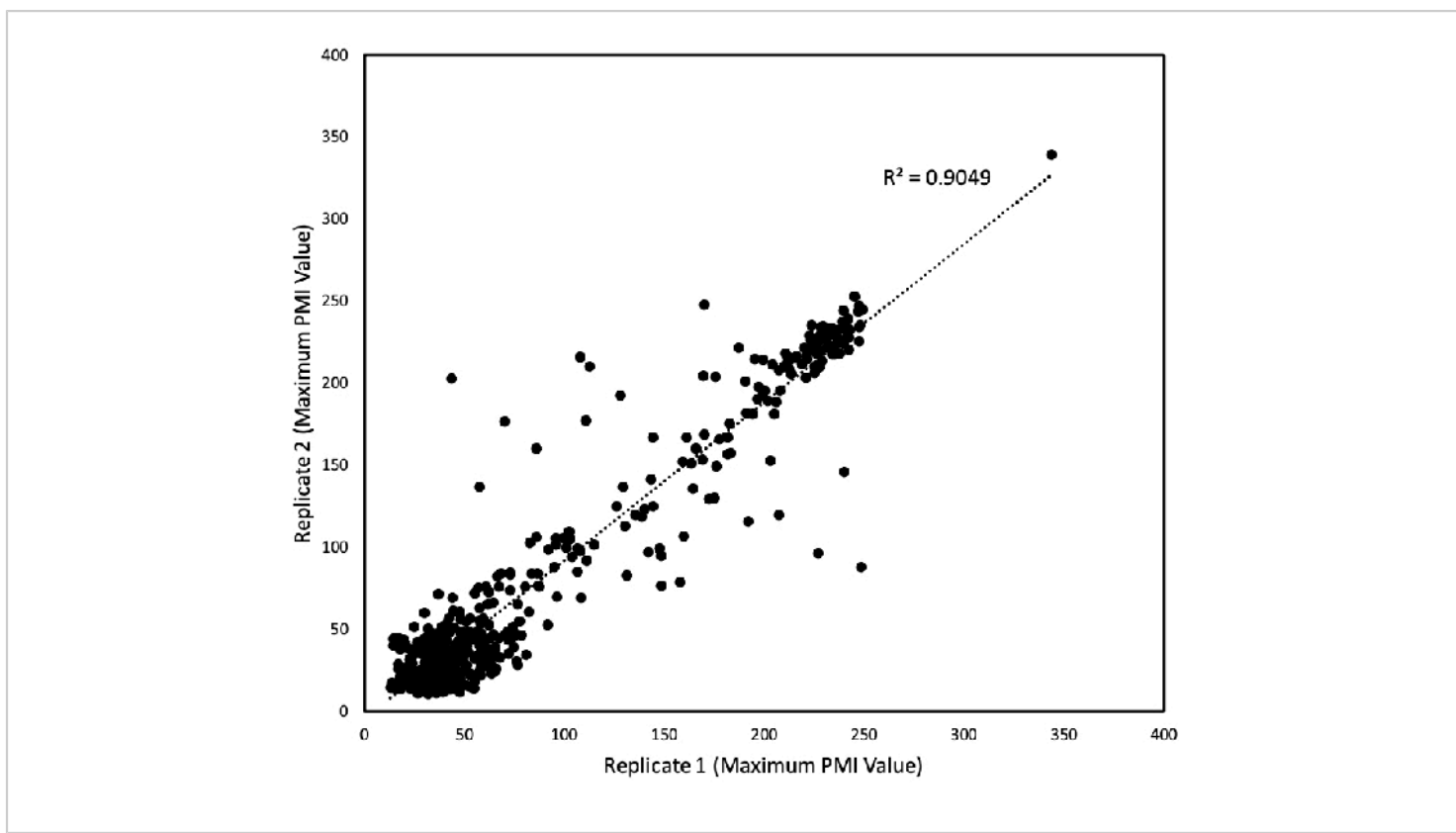


Figure 3: Reproducibility of PM tests. The PMI values were collected over a 168 hours period, and the maximum PMI values were plotted for two replicate studies. Each axis represents the maximum PMI values for each study (the x-axis being one replicate study and the y-axis another). Reproduced values are equidistant from each axis. Each point represents a single maximum value. The linear regression was performed by a spreadsheet software, and the resulting coefficient of determination (R^2) is shown. This figure has been modified from previously published work by Chaiboonchoe *et al.*¹² [Please click here to view a larger version of this figure.](#)

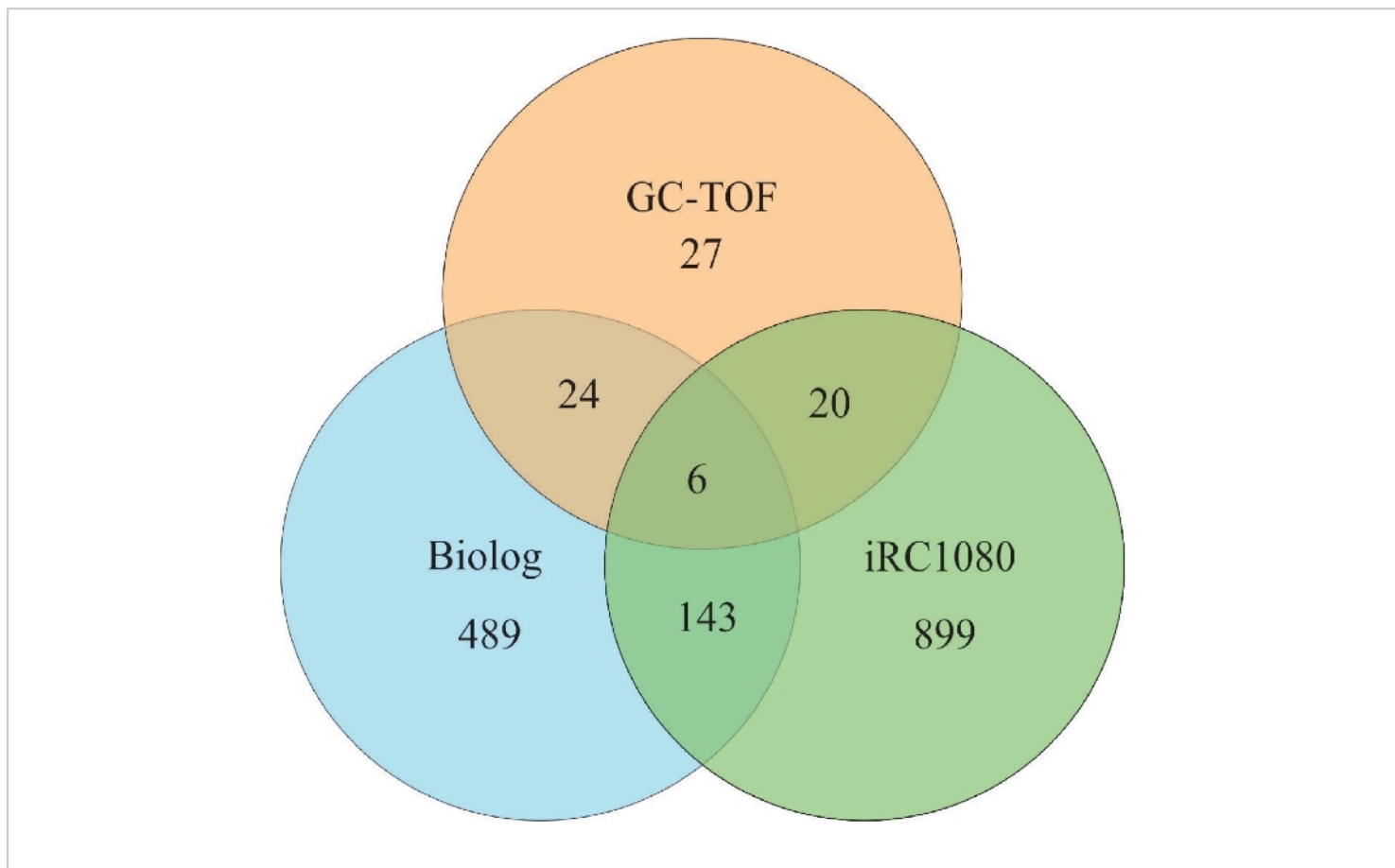


Figure 4: Venn diagram of metabolites. The Venn diagram enumerates metabolites identified by PM plates, the *iRC1080* metabolic model, and Gas Chromatography Time of Flight (GC-TOF) experiments. Each circle indicates the total number of metabolites that exist in each respective method of study. At the same time, the overlapping regions represent the number of metabolites shared between those methods. The *iRC1080* metabolic model contains a total of 1,068 unique metabolites. The GC-TOF identified a total of 77 metabolites³², while there are a total of 662 metabolites identified using the PM plates. This figure is from previously published work by Chaiboonchoe *et al.*¹² [Please click here to view a larger version of this figure.](#)

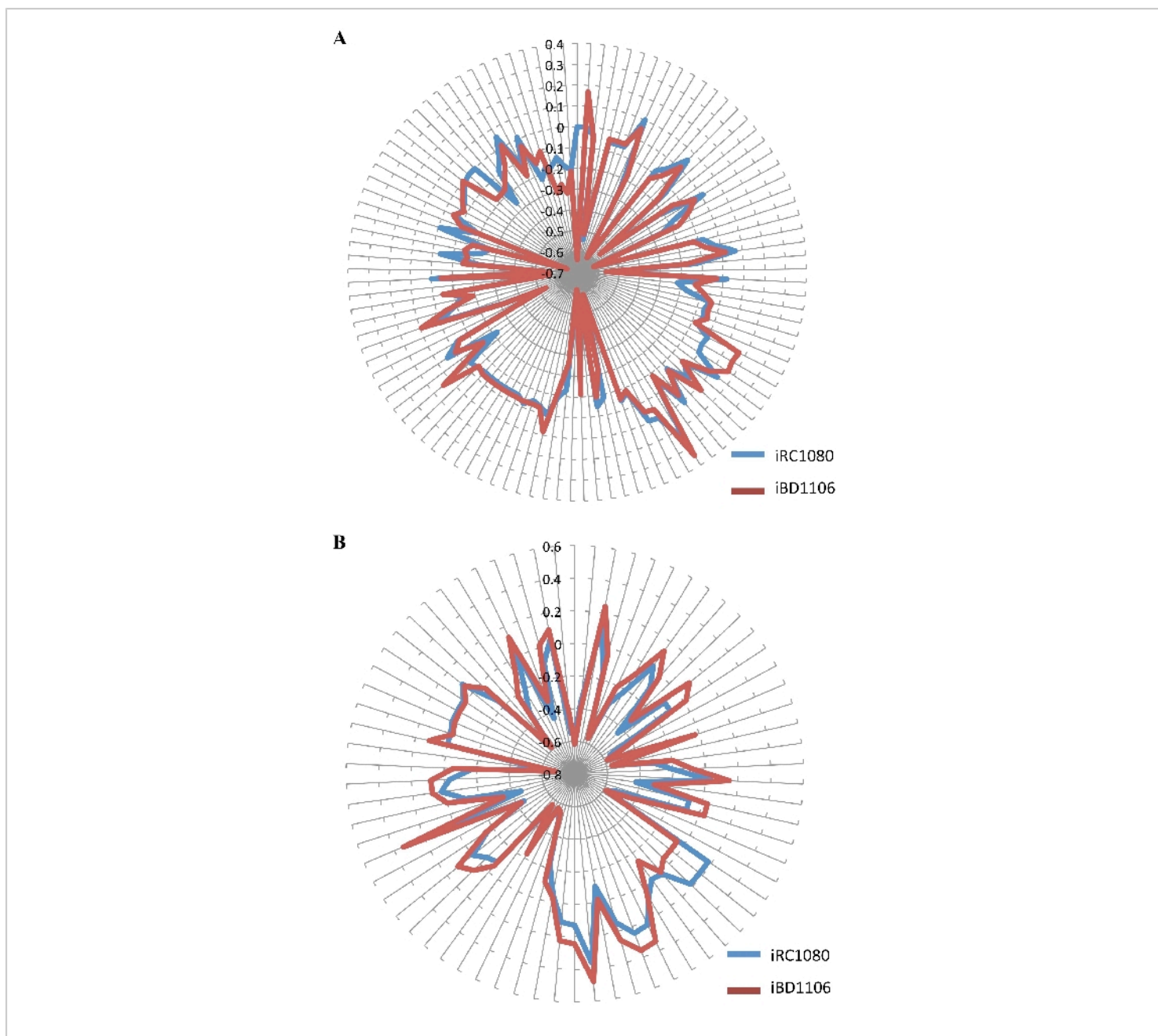


Figure 5: Shadow prices of metabolites in *iRC1080* and *iBD1106* under different conditions for biomass maximization. Each circle on the "radar plots" corresponds to a shadow price value, while each line extending from the center of a plot indicates a metabolite. **(A)** Shadow prices and metabolic behaviors of *iRC1080* and *iBD1106* under a light growth condition; **(B)**, different metabolic behaviors of *iRC1080* and *iBD1106* under a dark growth condition. This figure is from previously published work by Chaiboonchoe *et al.*¹² [Please click here to view a larger version of this figure.](#)

Biolog Chemical	EC*	Gene Annotation	PSI-BLAST
Cysteamine-S-Phosphate	3.1.3.1	JLM_162926 ^{1,2,3,4}	
Tetrathionate	1.8.2.2		insignificant E-value
	1.8.5.2		insignificant E-value
D-Alanine	1.4.1.1		XP_001700222.1
	1.5.1.22		failed manual QC
	2.1.2.7		insignificant E-value
	1.4.3.3	Cre02.g096350.t1.3 ⁵	
	2.3.2.10		insignificant E-value
	2.3.2.14		insignificant E-value
	2.3.2.16		insignificant E-value
	2.3.2.17		insignificant E-value
	2.3.2.18		insignificant E-value
	2.6.1.21		failed manual QC
	3.4.13.22		XP_001698572.1, XP_001693532.1, XP_001701890.1, XP_001700930.1
	3.4.16.4	Chlre2_kg.scaffold_ 14000039 ^{1,2,3}	
	3.4.17.8		failed manual QC
	3.4.17.13		insignificant E-value
	3.4.17.14		insignificant E-value
	4.5.1.2		insignificant E-value
	6.1.1.13		failed manual QC
	6.1.2.1		failed manual QC
	6.3.2.4	au.g14655_t1 ^{1,2,3}	
	6.3.2.10		failed manual QC

	6.3.2.16		insignificant E-value
	6.3.2.35		insignificant E-value
D-Asparagine	1.4.5.1		insignificant E-value
	1.4.3.3	Cre02.g096350.t1.3 ⁵	
	3.1.1.96		insignificant E-value
	2.3.1.36		insignificant E-value
	1.4.99.1		XP_001692123.1
	3.5.1.77	e_gwW.1.243.1 ^{1,2}	
	3.5.1.81		insignificant E-value
	5.1.1.10		failed manual QC
D-Aspartic Acid	6.3.1.12		insignificant E-value
	1.4.3.3	Cre02.g096350.t1.3 ⁵	
D-Glutamic Acid	1.4.3.7		insignificant E-value
	1.4.3.3		insignificant E-value
D-Lysine	5.4.3.4		insignificant E-value
	1.4.3.3	Cre02.g096350.t1.3 ⁵	
	6.3.2.37		failed manual QC
D-Serine	2.7.11.8		insignificant E-value
	2.7.11.17	Cre12.g486350.t1.3 ^{1,2,3,4}	
	3.4.21.78		failed manual QC
	3.4.21.104		failed manual QC
	4.3.1.18	g6244.t1 ⁴	failed manual QC
	6.3.2.35		insignificant E-value
	6.3.3.5		insignificant E-value
	1.4.3.3	Cre02.g096350.t1.3 ⁵	
D-Valine	1.21.3.1		failed manual QC
	6.3.2.26		failed manual QC

	1.4.3.3	Cre02.g096350.t1.3 ⁵	
L-Pyroglutamic Acid			
Thiophosphate			
Dithiophosphate			
Ethylamine	6.3.1.6		
D,L-a-Amino-Butyric Acid	2.1.1.49		insignificant E-value
	1.4.3.3	Cre02.g096350.t1.3 ⁵	
Di-peptide	3.4.13.18	Cre02.g078650.t1.3 ¹	
Tri-peptide	3.4.11.4	Cre16.g675350.t1.3 ¹	

Table 1: List of identified positive substrate utilization metabolites (C, P, S, N) not present in the *iRC1080* metabolic model. *Reaction was not included if no gene was identified. ¹Phytozome version 10.0.2 (http://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Creinhardtii). ²JGI version 4³⁵. ³Augustus version 5¹⁰. ⁴KEGG (<http://www.genome.jp/kegg/kegg1.html>). ⁵JGI version 3.1³⁶. This table is from previously published work by Chaiboonchoe *et al.* ¹²

Model	Reactions	Metabolites	Genes
<i>iRC1080</i>	2,191	1,706	1,086
<i>iBD1106</i>	2,445	1,959	1,106

Table 2: Contents of *iRC1080* and *iBD1106*. This table is from previously published work by Chaiboonchoe *et al.* ¹²

Category or Class of reactions	Number of reactions
Amino Acids	20
Dipeptides	108
Tripeptides	5
Transport reaction	120

Table 3: Summary of new reactions in *iBD1106*. This table is from previously published work by Chaiboonchoe *et al.* ¹²

Growth condition	Metabolite	Name	iRC1080	iBD1106
	4r5au	4-(1-D-Ribitylamino)-5-aminouracil	0	0.168
	5aprbu	5-Amino-6-(5'-phosphoribitylamino)uracil	-0.009	0.158
Light	pa1819Z18111Z	1-(9Z)-octadecenoyl,2-(11Z)-octadecenoyl-sn-glycerol3-phosphate	-0.009	-0.65
Dark	4abut	4-aminobutanoate	0.18	-0.05

Table 4: Example of significant shadow prices for iRC1080 and iBD1106. This table is from previously published work by Chaiboonchoe *et al.* ¹²

Discussion

Metabolic phenotyping of the green microalga, *C. reinhardtii*, was described here using high throughput PM assay plates and an unmodified PMI. The assays were utilized for a total of 190 carbon sources (PM01 and PM02), 95 nitrogen sources (PM03), 59 phosphorus sources, and 35 sulfur sources (PM04), along with peptide nitrogen sources (PM06-08). Positive respiration was observed for 148 nutrients (one positive assay for C-source utilization, four positive assays for each the S-source and P-source utilizations, and 139 positive assays for N-source utilization). The substrates or nutrients (carbon, nitrogen, phosphorus, or sulfur) component of the media should not be added to the defined medium when applied to the relevant PM microplates that test for each of those sources.

The method shown here is effective for characterizing metabolic microalgae phenotypes that can be used to extend existing metabolic network models or direct the reconstruction of new models. Further, as the nutritional requirements of most microalgae are not known, this platform can be used to define these rapidly. Nelson *et al.*⁴³ had successfully applied these methods to identify new compounds that support

the growth of the microalgae *Chloroidium* sp. UTEX 3007 and used the obtained information to define the species entry metabolites, which, unlike *Chlamydomonas*, include 40 different carbon sources.

One major limitation of the PM for profiling microalgae is that the PMI has no illumination in the incubation chamber, and the microalgae need to be able to carry out heterotrophic metabolism. The absence of light could affect the interpretation of models that incorporate light to calculate metabolic fluxes. Gene pairs with coordinating functions have co-evolved to constitute metabolic network hubs, and the distinction between photosynthetic and non-photosynthetic network hubs can be made⁴⁴. In general, photosynthetic network hubs (i.e., highly-connected nodes in the model) would be left out of heterotrophic models. For practical purposes, modeling heterotrophism in mixotrophic species should omit reactions known to be driven by light and account for the energy balance differences between conditions. Thus, modeling light-dependent and light-independent metabolism is standard practice in *Chlamydomonas* metabolic modeling^{6,45}.

Some green microalgae, like Trebouxiophytes, are known to assimilate a variety of carbon molecules for growth, and this is thought to have arisen from their long evolutionary history as members of lichens⁴⁶. While chlorophytes like *Chlamydomonas* can use acetate for growth, the brown marine microalga *Tisochrysis lutea*, known for its potential to commercially produce very-long-chain polyunsaturated fatty acids (VLC-PUFAs), cannot use acetate but can use glycerol for growth⁴⁷. Biomass concentration of more than 100 g l⁻¹ dry cell weight has been achieved with *Chlorella* with optimized addition of organic carbon sources in a fed-batch mode⁴⁸. Further, the addition of sugar to *Chlorellavulgaris* can elevate its sequestration of CO₂, thus providing an additive benefit during photosynthetic growth⁴⁹. Most heterotrophic microalgae can also grow mixotrophically, but the Chlorophyte *Chromochloris zofingiensis* has been shown to shut off photosynthesis upon the addition of sugar⁵⁰.

Diatoms, belonging to the division Bacillariophyta, are a major group of phytoplankton. Although most of the diatoms can only grow photoautotrophically, some of them can be cultivated mixotrophically or heterotrophically⁵¹. For example, glycerol was found to support growth in the light in the absence of CO₂ in some diatoms, including the model species *Phaeodactylum tricornutum*⁵². Also, some benthic diatoms like *Nitzschia linearis* can grow on carbohydrates in the dark⁵³. It is likely to extend the PM assays to diatoms and other algal groups by supplementing suitable organic carbon sources to enable the cells to grow heterotrophically, and a mixotrophy strategy can also be potentially used for the obligate autotrophic microalgae providing a minimally required light supply.

To assess the reproducibility of the data, it is highly recommended to carry out duplicate assays for all plates. An assay may be considered positive only if, after subtraction from the negative control and the respective blank wells, the absorbance (PMI value) is positive. This description, in the presence of the tested compound, is a reflection of the dye's abiotic reaction with the medium.

Disclosures

The authors have nothing to disclose.

Acknowledgments

Major support for this work was provided by the NYUAD Center for Genomics and Systems Biology (CGSB), funded by Tamkeen under New York University Abu Dhabi Research Institute grant (73 71210 CGSB9) and NYU Abu Dhabi Faculty Research Funds (AD060). W.F. was additionally supported by the Hundred Talents Program of Zhejiang University. We thank Ashish Jaiswal for help in recording the video. We thank Hong Cai for generating the metabolic phenotype data.

References

1. Oberhardt, M. A., Palsson, B. Ø., Papin, J. A. J. M. s. b. Applications of genome-scale metabolic reconstructions. *Molecular Systems Biology*. **5** (1), 320 (2009).
2. Schmidt, B. J., Lin-Schmidt, X., Chamberlin, A., Salehi-Ashtiani, K., Papin, J. A. Metabolic systems analysis to advance algal biotechnology. *Biotechnology Journal*. **5** (7), 660-670 (2010).
3. Koskimaki, J. E., Blazier, A. S., Clarens, A. F., Papin, J. A. J. I. B. Computational models of algae metabolism for industrial applications. *Industrial Biotechnology*. **9** (4), 185-195 (2013).

4. Koussa, J., Chaiboonchoe, A., Salehi-Ashtiani, K. J. B. *r. i.* Computational approaches for microalgal biofuel optimization: a review. *BioMed Research*. **2014**, 649453 (2014).
5. Nelson, D. R. *et al.* Large-scale genome sequencing reveals the driving forces of viruses in microalgal evolution. *Cell Host & Microbe*. **29** (2), 250-266. e258 (2021).
6. Shene, C., Asenjo, J. A., Chisti, Y. Metabolic modelling and simulation of the light and dark metabolism of *Chlamydomonas reinhardtii*. *The Plant Journal*. **96** (5), 1076-1088 (2018).
7. Tibocho-Bonilla, J. D., Zuñiga, C., Godoy-Silva, R. D., Zengler, K. Advances in metabolic modeling of oleaginous microalgae. *Biotechnology for Biofuels*. **11** (1), 241 (2018).
8. Merchant, S. S. *et al.* The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. *Science*. **318** (5848), 245-250 (2007).
9. May, P., Christian, J.-O., Kempa, S., Walther, D. J. B. G. ChlamyCyc: an integrative systems biology database and web-portal for *Chlamydomonas reinhardtii*. *BMC Genomics*. **10** (1), 209 (2009).
10. Chang, R. L. *et al.* Metabolic network reconstruction of *Chlamydomonas* offers insight into light-driven algal metabolism. *Molecular Systems Biology*. **7** (1) (2011).
11. de Oliveira Dal'Molin, C. G., Quek, L.-E., Palfreyman, R. W., Nielsen, L. K. AlgaGEM - a genome-scale metabolic reconstruction of algae based on the *Chlamydomonas reinhardtii* genome. *BMC Genomics*. **12** (S5) (2011).
12. Chaiboonchoe, A. *et al.* Microalgal metabolic network model refinement through high-throughput functional metabolic profiling. *Frontiers in Bioengineering and Biotechnology*. **2**, 68 (2014).
13. Kanehisa, M. *et al.* Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Research*. **42** (D1), D199-D205 (2014).
14. Zuñiga, C. *et al.* Genome-scale metabolic model for the green alga *Chlorella vulgaris* UTEX 395 accurately predicts phenotypes under autotrophic, heterotrophic, and mixotrophic growth conditions. *Plant Physiology*. **172** (1), 589-602 (2016).
15. Bochner, B. R. New technologies to assess genotype-phenotype relationships. *Nature Reviews Genetics*. **4** (4), 309-314 (2003).
16. Bochner, B. R. Global phenotypic characterization of bacteria. *FEMS Microbiology Reviews*. **33** (1), 191-205 (2009).
17. Bochner, B. R., Gadzinski, P., Panomitros, E. Phenotype microarrays for high-throughput phenotypic testing and assay of gene function. *Genome Research*. **11** (7), 1246-1255 (2001).
18. Bartell, J. A., Yen, P., Varga, J. J., Goldberg, J. B., Papin, J. A. Comparative metabolic systems analysis of pathogenic *Burkholderia*. *Journal of Bacteriology*. **196** (2), 210-226 (2014).
19. Gorman, D. S., Levine, R. J. P. o. t. N. A. o. S. Cytochrome f and plastocyanin: their sequence in the photosynthetic electron transport chain of *Chlamydomonas reinhardtii*. *PNAS*. **54** (6), 1665-1669 (1965).
20. Smith, A. C., Hussey, M. A. Gram stain protocols. *American Society for Microbiology*. 1-9 (2005).

21. Vaas, L. A. I. *et al.* *opm*: an R package for analysing OmniLog phenotype microarray data. *Bioinformatics*. **29** (14), 1823-1824 (2013).
22. Vaas, L. A. I., Sikorski, J., Michael, V., Göker, M., Klenk, H.-P. Visualization and Curve-Parameter Estimation Strategies for Efficient Exploration of Phenotype Microarray Kinetics. *PLoS ONE*. **7** (4), e34846 (2012).
23. Caspi, R. *et al.* The MetaCyc database of metabolic pathways and enzymes—a 2019 update. *Nucleic Acids Research*. **48** (D1), D445-D453 (2020).
24. Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M., Tanabe, M. KEGG: integrating viruses and cellular organisms. *Nucleic Acids Research*. (2020).
25. Lopez, D., Casero, D., Cokus, S. J., Merchant, S. S., Pellegrini, M. Algal Functional Annotation Tool: a web-based analysis suite to functionally interpret large gene lists using integrated annotation and expression data. *BMC Bioinformatics*. **12** (1), 282 (2011).
26. Caspi, R. *et al.* The MetaCyc database of metabolic pathways and enzymes. *Nucleic Acids Research*. **46** (D1), 633-639 (2018).
27. Sahoo, S. *et al.* dEMBF v2. 0: An Updated Database of Enzymes for Microalgal Biofuel Feedstock. *Plant and Cell Physiology*. **61** (5), 1019-1024 (2020).
28. Heirendt, L. *et al.* Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v. 3.0. *Nature Protocols*. **14** (3), 639-702 (2019).
29. Heirendt, L. *et al.* Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v. 3.0. *Nature Protocols*. **1** (2019).
30. Orth, J. D., Thiele, I., Palsson, B. Ø. What is flux balance analysis? *Nature Biotechnology*. **28** (3), 245 (2010).
31. Ebrahim, A., Lerman, J. A., Palsson, B. O., Hyduke, D. R. COBRApy: constraints-based reconstruction and analysis for python. *BMC Systems Biology*. **7** (1), 74 (2013).
32. Bölling, C., Fiehn, O. Metabolite profiling of *Chlamydomonas reinhardtii* under nutrient deprivation. *Plant Physiology*. **139** (4), 1995-2005 (2005).
33. Harris, E. H. *The Chlamydomonas sourcebook: introduction to Chlamydomonas and its laboratory use*. Vol. 1 Academic Press, (2009).
34. Goodstein, D. M. *et al.* Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Research*. **40** (D1), D1178-D1186 (2012).
35. Ghamsari, L. *et al.* Genome-wide functional annotation and structural verification of metabolic ORFeome of *Chlamydomonas reinhardtii*. *BMC Genomics*. **12** (1), S4 (2011).
36. Manichaikul, A. *et al.* Metabolic network analysis integrated with transcript verification for sequenced genomes. *Nature Methods*. **6** (8), 589-592 (2009).
37. Apweiler, R. *et al.* UniProt: the Universal Protein knowledgebase. *Nucleic Acids Research*. **32** (Database issue), D115-119 (2004).
38. Consortium, T. U. Activities at the universal protein resource (UniProt). *Nucleic Acids Research*. **42** (11), 7486-7486 (2014).
39. Horton, P. *et al.* WoLF PSORT: protein localization predictor. *Nucleic Acids Research*. **35** (suppl_2), W585-W587 (2007).
40. Becker, S. A. *et al.* Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. *Nature Protocols*. **2** (3), 727-738 (2007).

41. Schellenberger, J. *et al.* Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2. 0. *Nature Protocols*. **6** (9), 1290 (2011).
42. Varma, A., Boesch, B. W., Palsson, B. O. Stoichiometric interpretation of Escherichia coli glucose catabolism under various oxygenation rates. *Applied and Environmental Microbiology*. **59** (8), 2465-2473 (1993).
43. Nelson, D. R. *et al.* The genome and phenome of the green alga Chloroidium sp. UTEX 3007 reveal adaptive traits for desert acclimatization. *eLife*.10.7554/eLife.25783 (2017).
44. Chaiboonchoe, A. *et al.* Systems level analysis of the Chlamydomonas reinhardtii metabolic network reveals variability in evolutionary co-conservation. *Molecular BioSystems*. **12** (8), 2394-2407 (2016).
45. Chang, R. L. *et al.* Metabolic network reconstruction of Chlamydomonas offers insight into light-driven algal metabolism. *Molecular Systems Biology*. **7** (1), 518 (2011).
46. Rajendran, A., Hu, B. Mycoalgae biofilm: development of a novel platform technology using algae and fungal cultures. *Biotechnology for Biofuels*. **9** (1), 112 (2016).
47. Hu, H. *et al.* Effect of cultivation mode on the production of docosahexaenoic acid by Tisochrysis lutea. *AMB Express*. **8** (1), 50 (2018).
48. Bumbak, F., Cook, S., Zachleder, V., Hauser, S., Kovar, K. Best practices in heterotrophic high-cell-density microalgal processes: achievements, potential and possible limitations. *Applied Microbiology and Biotechnology*. **91** (1), 31 (2011).
49. Fu, W. *et al.* Sugar-stimulated CO₂ sequestration by the green microalga Chlorella vulgaris. *Science of the Total Environment*. **654**, 275-283 (2019).
50. Roth, M. S. *et al.* Regulation of oxygenic photosynthesis during trophic transitions in the green alga Chromochloris zofingiensis. *The Plant Cell* .(2019).
51. Villanova, V. *et al.* Investigating mixotrophic metabolism in the model diatom Phaeodactylum tricornutum. *Philosophical Transactions of the Royal Society B: Biological Sciences*. **372** (1728), 20160404 (2017).
52. Cerón-García, M. *et al.* Mixotrophic growth of Phaeodactylum tricornutum on fructose and glycerol in fed-batch and semi-continuous modes. *Bioresource Technology*. **147**, 569-576 (2013).
53. Tuchman, N. C., Schollett, M. A., Rier, S. T., Geddes, P. in *Advances in Algal Biology: A Commemoration of the Work of Rex Lowe*.167-177 Springer, (2006).